Development of nanoflow liquid chromatography-nanoelectrospray ionization mass spectrometry methodology for improved urine metabolomics

Andrew John Chetwynd

PhD

University of Sussex

Submitted March 2015

Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:..... Andrew Chetwynd

Acknowledgements

There are many people to whom I owe many thanks for their support, advice and help throughout my PhD. First of all I wish to thank both of my supervisors Elizabeth Hill and Alaa Abdul-Sada for the opportunity to complete my PhD in their group. I am very grateful for their knowledge, guidance and support for the entirety of my study.

I am particularly thankful to Paolo Indiveri and Raghad Al-Salhi for their encouragement and valuable advice when I first started my thesis. Extra thanks go to Arthur David who has patiently provided answers to countless questions for the past few years and given expert advice on a wide range of topics. Huge thanks also go to Julia for her knowledge and guidance on the human tissue act and for ensuring the lab ran smoothly. I also owe huge thanks to Inma Jiménez-Díaz whose friendship, kind words of advice and encouragement helped make the final year of my PhD less stressful and more enjoyable.

I also wish to thank the clinicians, in particular Dr Amanda Samarawickrama, at Brighton and Sussex University Hospital who collected the samples used in the case studies within this thesis. Further thanks go to Gary Weaving from the pathology lab at Brighton and Sussex University Hospital for allowing me access to some of their equipment and for his advice and support while I worked in the pathology lab.

I wish to offer a big thanks to the entire Hill group with whom I have worked with in the past few years: Pawel Rostkowski, Paolo Indiveri, Camilla Liscio, Raghad Al-Salhi, Arthur David, Diana Alvarez-Muñoz, Inmaculada Jiménez-Díaz, Julia Horwood, Cristina Botías-Talamantes, Cathy Ma and Emily Greer. Your friendships and support made completing my PhD a more enjoyable experience.

Finally I wish to thank my parents, brother and grandparents whose unending care, support and encouragement for the past 26 years made all of this possible in the first place.

UNIVERSITY OF SUSSEX

ANDREW JOHN CHETWYND

PhD

Development of nanoflow liquid chromatography-nanoelectrospray ionization mass spectrometry methodology for improved urine metabolomics

SUMMARY

Global metabolomic analysis of urine offers great potential for detection of early warning markers of disease. Current methods focus on rapid sample preparation and high throughput analyses at the expense of the detection of low abundance metabolites. The aim of this study was to develop sensitive analytical methods for metabolomic profiling. Methods were developed to use nanoflow ultra high performance liquid chromatography-nanospray ionization-mass spectrometry (nUHPLC-nESI-TOFMS), normally used for proteomics, for metabolomic analyses of urine samples. Compared with a conventional UHPLC-ESI-TOFMS, the use of a nanoflow-nanospray platform increased the sensitivity to a standard mixture of metabolites by 2-2000 fold. Highly repeatable results for retention time and metabolome peak area were achieved, where the coefficients of variation were <0.2% and <30% respectively for the majority of peaks present in the urine metabolome. To further increase sensitivity and enable small injection volumes, a sample preparation method was developed using polymeric anion and cation exchange mixed mode solid phase extraction with pre-concentration. Combined with the nano platform, this enabled the detection of low abundance signalling molecules (estrogens, eicosanoids and unconjugated androgens) not usually detected with conventional methods. A pre-analysis normalisation technique based on osmolality concentrations was used to reduce sample variability due to differing urine concentrations. These methods were used to investigate the metabolomic consequences of HIV infection and patient response to combined antiretroviral therapy (cART). No significant differences in metabolomic profiles between HIV positive and negative patients were observed. However, disruption of bile acid profiles and decreased concentrations of selected carnitines, steroid conjugates, polypeptides and nucleosides were detected in patients on cART therapy indicating disrupted lipid and protein metabolism but improved immunological function associated with antiretroviral medication. These finding highlight the importance of these newly developed SPE sample preparation and nUHPLC-nESI-TOFMS analysis methods for global profiling of the urinary metabolome.

Publications

To date, two research papers have been published as a result of the work presented in this thesis and these papers form Chapters 2 and 3. Full details for these publications are provided below.

Chetwynd, A.J.; David, A.; Hill, E.M.; Abdul-Sada, A. Evaluation of sensitivity and reliability of direct nanoLC-nanoESI-high resolution mass spectrometry for metabolomic profiling. Journal of Mass Spectrometry, 2014, 49, 1063-1069.

Chetwynd, A.J.; Abdul-Sada, A.; Hill, E.M. Solid phase extraction and nanoflow liquid chromatography-nanoelectrospray ionization mass spectrometry for improved global urine metabolomics, Analytical Chemistry, 2015, 87, 1158-1165.

Table of Contents

List of figures and tables

Chapter 1

Abbreviations

%CV	Percentage covariance
+ve	Positive
2D	2 dimensional
AA	Arachidonic acid
ABC	ATP binding cassette transporter
ACN	Acetonitrile
AIDS	Acquired immunodeficiency disease
AKI	Acute kidney injury
AN	Androstenedione
ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photoionization
ART	Antiretroviral therapy
ARV	Antiretroviral
atz	Atazanavir
AUC	Area under the curve
BEH	Ethylene Bridged Hybrid
BPI	Base peak intensity
BSEP	Bile Salt Export Pump
cART	Combined antiretroviral therapy
CCS	Collisional cross section
CI	Chemical ionization
CSF	Cerebrospinal fluid
Cyp	Cytochrome
d.c	Direct current
DIMS	Direct infusion mass spectrometry

DNA	Deoxyribonucleic acid
drn	Darunavir
EI	Electron ionization
ESI	Electrospray ionization
Exp.	Experimental
FA	Formic acid
FT-ICR	Fourier transformer-ion cyclotron resonance
GC-MS	Gas chromatography mass spectrometry
gp	Glycoprotein
H_2O	Water
HETE	Hydroxyeicosatetraenoic acids
HILIC	Hydrophobic interaction liquid chromatography
HIV	Human immunodeficiency virus
HMDB	Human metabolome database
HPLC	High pressure liquid chromatography
HSS	High Strength Silica
IBP	Ibuprofen
id	Internal diameter
IDL	Instrument detection limit
IM	Ion mobility
IS	Internal standard
LC-MS	Liquid chromatography mass spectrometry
LogKow	Octanol-water partition coefficient
LT	Leukotriene
m/z	Mass to charge ration
M+H	Protonated mass

MALDI	Matrix assisted laser desorption ionization
MeOH	Methanol
M-H	Deprotonated mass
MS	Mass spectrometer
MSTS	Mass spectrometer total signal
MSTUS	Mass spectrometer total useful signal
MVA	Multivariate analysis
nd	Not detected
nESI	Nanospray electrospray ionization
nLC	Nanoflow liquid chromatography
NMR	Nuclear Magnetic Resonance
NNRTI	Non nucleotide reverse transcriptase inhibitor
NRTI	Nucleotide reverse transcriptase inhibitor
NSAID	Non-steroidal anti inflammatory drug
nUHPLC	Nanoflow ultra high performance liquid chromatography
OATP	Organic anion transporting polypeptides
OPLS-DA	Orthogonal partial least square discriminant analysis
P	Progesterone
PCA	Principal component analysis
PEEK	Polyetheretherketone
PI	Protease inhibitor
PLS-DA	Partial least square discriminant analysis
ppm	Parts per million
PR	Propranolol
psi	Pounds per square inch
PW	Peak width

QC	Quality control
QTOF	Quadrupole time of flight
r.f	Radio frequency
RNA	Ribonucleic acid
ROC	Receiver operator curve
RP	Reversed phase
RPLC	Reversed phase liquid chromatography
Rt	Retention time
S/N	Signal to noise ration
SD	Standard deviation
SIM	Selected ion monitoring
SPE	Solid phase extraction
SSRI	Selective serotonin reuptake inhibitor
T2	Testosterone
T2G	Testosterone glucuronide
TBEP	Tris(2-butoxyethyl)phosphate
TCA	Taurocholic acid
Thr.	Theoretical
TOF	Time of flight
TxB2	Thromboxane B2
UHPLC	Ultra high pressure liquid chromatography
UPCR	Urinary protein creatinine ratio
-ve	Negative

Chapter 1: General Introduction

1.1 The -Omics sciences

The -omic sciences are a rapidly expanding area of analytical approaches which generate vast amounts of experimental data. Through analysis of these data mechanisms of disease, physiology and toxic exposures can be elucidated (Alvarez-Sanchez et al., 2010a). The 4 major –omics are genomics the study of an organism's genome (Sadee, 2011), proteomics the study of the proteins within an organism (Perrett, 2007), transcriptomics which studies the RNA transcripts of the genome (Hegde et al., 2003) and metabolomics the study of an organism's metabolites (Nicholson et al., 1999). The first three approaches study products that are entirely dependent on the nature of the encoded genome, whereas metabolomics is concerned with the products of both physiological and cellular metabolism and therefore more closely reflects the phenotype (Fiehn, 2002, Vinayavekhin et al., 2010). The combination of -omic sciences provide models and uncover emergent properties of an entire organism which can be split to its constituent organs/biofluids and subsequent cells, all falling under the remit of systems biology (Nicholson and Wilson, 2003, Bruggeman and Westerhoff, 2007, Dunn et al., 2011).

1.2 Metabolomics

In 1999 Nicholson coined the term to describe the analysis of small molecules in a biological matrix in response to pathological stimulus (Nicholson et al., 1999). Metabolomics is the semi quantitative study of the small molecules involved in enzymatic metabolic reactions, as either the reactants, intermediaries or products of such reactions (Dunn et al., 2011, Alvarez-Sanchez et al., 2010b). Since their inception, both metabolomics and metabonomics have been used interchangeably. Metabolites are defined as low molecular weight organic molecules typically less than 1000Da (Nicholson, 2006, Guy et al., 2008), and the metabolite collective of a sample matrix is known as the metabolome (Nicholson, 2006). Metabolomic variation between sample groups may be indicative of physiological stress or environmental exposures. The metabolites driving the metabolic variation under these stressors, once extensively validated, are often referred to as biomarkers (Want et al., 2010, Gika et al., 2014). Like the other major –omics sciences, with the exception of genomics, metabolomics is capable of being implemented across all levels of complexity within an organism, with metabolic profiles being possible for cells, tissues, biological fluids, cellular and sub-cellular compartments (Fiehn, 2002). In addition, metabolomics is also referred to as a top-down approach (Spagou et al., 2011, Nicholson and Lindon, 2008) whereby disrupted metabolic pathways are identified and then linked to changes in gene expression controlling these

pathways thus providing a detailed overview of the metabolic stress being exerted (Gika et al., 2014).

Metabolomics as we know it today is a relatively new area of science which only recently has been made possible by the advances in analytical techniques allowing hundreds if not thousands of molecules to be separated and identified with relative ease (Mamas et al., 2011). There are two strategies for metabolomic analysis for any given sample matrix, these are; non-targeted/global metabolomics and targeted metabolomics. The desired approach is determined based upon the aim of the study. Non-targeted or global metabolomics is the study of the entire metabolome of a biological sample and requires a methodology that allows for the detection of the greatest number of metabolites as possible (Issaq et al., 2009, Vinayavekhin et al., 2010, Raterink et al., 2014). This method is a discovery-based approach where there is limited prior knowledge of the biology involved in the case study. Comparison of case study metabolomes with matched controls utilizing multivariate statistics enables the identification of discriminating metabolites as the result of metabolic challenges. Global metabolomics is generally a theory producing experiment, where variations in metabolome profiles are used to produce theories about the cause of the metabolite variation and how it influences disease state (Mamas et al., 2011). In contrast, a targeted metabolomics approach is a (semi)quantitative analysis of a small number of metabolites which are usually known in advance (Vinayavekhin et al., 2010, Dunn et al., 2011, Issaq et al., 2009). The biology behind them is well understood, possibly due to a previous global analysis (Mamas et al., 2011). This method is typically used once a potential biomarker or collection of potential biomarkers has already been identified, and it becomes necessary to (semi)quantify these metabolites to diagnose or observe a disease state.

1.3 Analytical platforms for metabolomic analysis

The diverse range of metabolites within the metabolome from any given matrix offers significant challenges for analytical chemists to overcome. The human metabolome comprises more than 6,500 characterised metabolites (Wishart et al., 2009, Wishart et al., 2007, Wishart et al., 2013) a number which will only grow with new improved analytical platforms. In addition to endogenously derived metabolites, there are exogenously derived metabolites often referred to as xenobiotics and xenometabolites (Holmes et al., 2007), which are introduced to the overall metabolome through the diet (Andersen et al., 2013, Pujos-Guillot et al., 2013, Wang et al., 2012a, Jacobs et al., 2012, Hall et al., 2011b, Edmands et al., 2011, Lloyd et al., 2011, Fave et al., 2009), pharmaceuticals (Loo et al., 2012, Holmes et al., 2007) and environmental exposure (Nomura et al., 2013, Ueyama et al., 2012, Johnson et al., 2012a).

When combined with the endogenous metabolites it is thought that the total number of metabolites present in the human metabolome could be as many as 40,000 (Wishart et al., 2009, Wishart et al., 2013). The entire human metabolome covers a broad range of polarities (Want et al., 2010) and spans at least 11 orders of magnitude in compound concentrations (Bouatra et al., 2013, Psychogios et al., 2011, Wishart et al., 2009, Wishart et al., 2007). With recent advancements in analytical chemistry there are a range of analytical platforms that can be implemented for metabolomic analysis. The commonly utilized techniques are spectroscopic techniques alone such as Nuclear Magnetic Resonance (NMR) and direct infusion mass spectrometry (DIMS), or hyphenated techniques incorporating chromatographic separation such as Gas Chromatography-Mass Spectrometry (GC-MS) and Liquid Chromatography-Mass Spectrometry (LC-MS) While no one analytical platform offers complete metabolome analysis (Issaq et al., 2009, Ramautar et al., 2009) a combination is deemed complementary, with each technique adding metabolome information that others cannot (Wilson et al., 2005, Raterink et al., 2014).

1.3.1 Detection techniques

The method utilised for the detection of metabolites plays an important role in the metabolomic study. The technique chosen will determine the dynamic range of the analysis and the ability to detect and identify metabolite types. Metabolomic analysis is dominated by two detection techniques, NMR and mass spectrometry (Zhang et al., 2012a).

1.3.1.1 NMR

Many of the initial metabolomics studies were carried out using ^{1}H or ^{13}C NMR (Emwas et al., 2013, Nicholson et al., 1999) and this technique is still used extensively today (Kaddurah-Daouk et al., 2008, Theodoridis et al., 2012). NMR analysis offers a quick non-destructive technique for metabolomic analysis with the additional benefit of structural elucidation of unknown metabolites (Lenz and Wilson, 2007, Emwas et al., 2013). Indeed structural elucidation of unknown metabolites and platform resolution has been significantly improved with the introduction of 2D NMR due to the reduction in overlapping peaks. This is achieved by collecting a series of 1D spectra and plotting data on a two frequency axis as opposed to one (Bingol and Bruschweiler, 2014, Le Guennec et al., 2014). There has been much success in NMR-led metabolomics, for example in environmental research (Viant et al., 2003, Samuelsson et al., 2006), human health and disease (Fischer et al., 2014, Serkova et al., 2007, Holmes et al., 2008) and drug discovery (Powers, 2009). Despite the advantages of structural elucidation offered for metabolomic analysis, LC-NMR has thus far not been widely adopted by the metabolomics community, possibly as a result of the reduced sensitivity compared to mass

spectrometry (Brennan, 2014, Albert, 1995). In addition, most LC runs metabolomics utilize a gradient system utilizing protonated solvents, typically water and acetonitrile. This creates changes in the solvent spectra over the course of the run meaning a more complex solvent suppression methodology is required. Another option would be to use fully deuterated solvents which would significantly increase the cost of analysis due to the volume of solvent required and the cost of deuterated solvents (Albert, 1995, Exarchou et al., 2005). However, NMR or LC-NMR metabolomics analysis suffers from a key limitation; limited sensitivity (Fischer et al., 2014, Emwas et al., 2013, Ramautar et al., 2009). However, the lack of sensitivity can be alleviated to some extent by using a larger magnetic field instrument, in addition to the miniaturisation of the NMR probes (Moco et al., 2007).

1.3.1.2 Mass spectrometry

The most commonly used detector in metabolomics analysis is mass spectroscopy (Scalbert et al., 2009) and as such a more detailed discussion of the technique is given below. Unlike NMR, MS is a destructive analytical platform thus resulting in loss of the sample and leaving only fragmentation data for structural elucidation of unknown compounds. The aim of mass spectrometry is the measurement of the mass to charge ratio of ions and their quantification (Guilhaus, 1995). There is a wide array of techniques available to achieve this using different methods of generating ions and different mass analysis techniques.

1.3.1.2.1 Ionization techniques

In order for a molecule to enter a mass analyser and be “directed” to the detector it must first be converted to a gaseous phase if not already so (de Hoffmann and Stroobart, 2007). In addition, an electrostatic charge is imparted upon analytes giving them either an overall positive or negative charge (de Hoffmann and Stroobart, 2007). The use of reflector plates within a mass spectrometer utilize electrostatic repulsion to remove unionized molecules and guide the ionized ones to the mass analyser and subsequent detector. A number of ionization techniques are available and can be divided into categories in two ways; the first using either under high vacuum (electron ionization and chemical ionization) or atmospheric pressure (electrospray ionization) induced ionization, the second by influencing the amount of fragmentation that occurs during ionization using techniques. Hard ionization creates plentiful fragments which can be used to elucidate structural information of the molecule; soft ionization techniques limit fragmentation meaning fewer ions are produced for a single molecule, which reduces the complexity of mass spectral data (de Hoffmann and Stroobart, 2007).

1.3.1.2.2 Electron ionization

Electron ionization (EI) is widely used in the field of GC-MS; electrons are emitted via thermic emission from a heated filament (de Hoffmann and Stroobart, 2007). These electrons are accelerated, within an electric field, typically to 70eV, between the filament and the source where the analytes enter the mass spectrometer (Bleakney, 1929). Interactions between the high energy electrons and gaseous analyte induce ionization, and due to the electrons having a de Broglie wavelength similar to that of the average bond length in organic molecules, fragmentation occurs resulting in a hard ionization technique. As a result of the hard ionization there is a high degree of fragmentation, meaning structural elements of the metabolite can be elucidated in addition to serving as a method of identification by analysing fragmentation patterns with spectral libraries. At its most efficient, only around 1 out of every 1000 molecules become ionized and thus enter the mass analyser, with negative mode ionization being much less efficient (de Hoffmann and Stroobart, 2007).

1.3.1.2.3 Chemical ionization

First introduced in 1966 chemical ionization (CI) offers a much less energetic ionization process compared to electron ionization, this means that much fewer fragments are produced in the process (Munson and Field, 1966). Here gaseous analyte is ionized via collisions with ions of a reagent gas such as ammonia. The reagent gas itself is ionized via EI, however, collisions between analyte and reagent gas are not very energetic meaning very few fragments are formed. This ensures that the molecular ion is easily identifiable and is the only protonated ion created in the process. Most fragments will not be protonated and present as a neutral mass (Harrison, 1992, de Hoffmann and Stroobart, 2007).

1.3.1.2.4 Matrix assisted laser desorption ionization (MALDI)

MALDI is typically used for large biomolecules; as such it has been extensively used in the field of proteomics. Analytes are dried to produce a solid matrix which is ablated by a high power laser, quickly heating the sample causing localised sublimation allowing the sample to enter the gas phase where it is protonated or deprotonated in the hot plume generated by the laser ablation. Ions subsequently enter the mass analyser. The soft ionization, and ability to produce multiply charged ions whilst not being detrimentally affected by salts and buffer solution, makes it ideal for large biomolecules applications such as proteomics, DNA and polysaccharides (de Hoffmann and Stroobart, 2007). MALDI however suffers from a relatively high level of noise at m/z values below 500Da due to the sample matrix, effectively making half the metabolome problematic to analyse (Glish and Vachet, 2003).

1.3.1.2.5 Atmospheric pressure ionization

The previously discussed ionization methods are ionized analytes under high vacuum; however several methods of ionization are available under atmospheric pressure. In addition, instead of ionization in the gas phase, analytes can be introduced to the source in an aqueous phase. Whereas in EI and CI analytes must be introduced in gas phase, and as such non-volatile compounds require derivatization and need to be thermo-stable. MALDI ionization requires the analytes to be bound to a solid matrix prior to analysis. The ability to analyse samples in liquid form is ideal for biological matrices as they are typically highly aqueous. This also eliminates the need for chemical derivatization or the formation of a solid matrix. There are several atmospheric pressure ionization techniques including electrospray ionization (ESI) atmospheric pressure chemical ionization (APCI) and atmospheric pressure photoionization (APPI) which will be discussed in the upcoming section (de Hoffmann and Stroobart, 2007).

1.3.1.2.5.1 Electrospray ionization

Invented in the late 1980s by Fenn, electrospray ionization (ESI) (Yamashita and Fenn, 1984) has become one of the most widely used ionization techniques for analytical chemists (Bruins, 1998). Initially used for large molecule analysis such as proteins and peptides due to its soft ionization at atmospheric pressure and ability to form multiply charged ions (Glish and Vachet, 2003, de Hoffmann and Stroobart, 2007, Fenn et al., 1989), ESI is now a popular method of ionization in metabolomic analysis (Want et al., 2010) Within the ESI source, the analyte is dispersed from the ESI probe and forms a Taylor cone due to the pneumatic drive from the LC or sample injector. The spray emitted is in the form of charged spherical droplets, and the charge is imparted by a large potential difference typically between 3 and 6 kV. The droplets undergo an evaporative process known as desolvation with the aid of a desolvation gas and a desolvation temperature (circa 300°C) (Figure 1.1). The combination of gas and temperature slowly shrinks the droplets by evaporating the mobile phase. Once the droplets reach their Rayleigh point the electrostatic charge overcomes the surface tension of the droplet dispersing it into smaller droplets in a coulomb explosion (Bruins, 1998, Smith et al., 1990, Fenn et al., 1989). Each of these droplets undergo further desolvation and coulomb explosion cycles before gas phase ions are formed which enter the mass analyser (Glish and Vachet, 2003, Juraschek et al., 1999). The efficiency of ionization is dependent upon the initial droplet size and this relationship between droplet radius and volume is described by Wickremsinhe et al., (Wickremsinhe et al., 2006). The smaller the initial droplet size the greater the rate of desolvation and this allows fewer desolvation cycles before an analyte can enter the gas phase and subsequently enter the mass analyser.

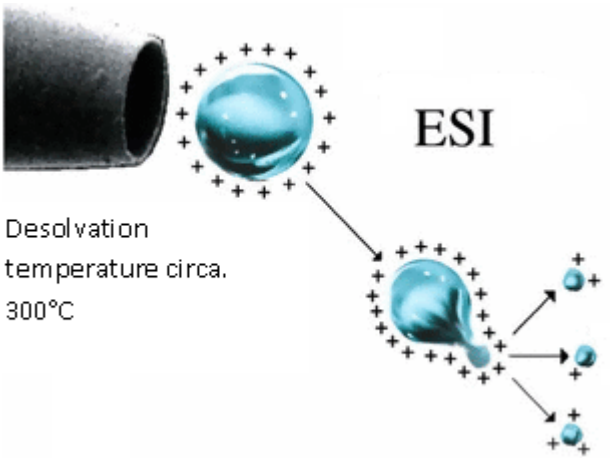


Figure 1.1: Mechanism of electrospray ionization
(Adapted from: "Biomarkers for NeuroAIDS: The Widening Scope of Metabolomics" 2007) (Pendyala et al., 2007)

In metabolomics when ESI is used as a hyphenated technique coupled with liquid-chromatography (LC), the role of mobile phase and additive is an important consideration for ionization. The protonation of a base to form a positively charge ion can be encouraged with the addition of a proton donor such as an acid to the mobile phase or using methanol as the organic solvent. Conversely negative ions are formed following reversal of the electrical potential of the capillary and by minimising the amount of acid present in the mobile phase or with the addition of basic additive such as ammonium hydroxide which sequesters a proton from the acidic molecule leaving a negative charge (Bruins, 1998).

1.3.1.2.5.1.1 ESI artefacts

A number of reactions can occur within an ESI source which alters the *m/z* of the analyte. When introduced at a high concentration analytes can form protonated dimers, trimers, etc, though these are more common in positive ESI mode. This is due to the greater stability of ions generated in positive ESI relative to negative ESI, meaning a greater proportion travel to the detector intact to be detected as a dimer/trimer ion (Ding and Anderegg, 1995). Other common ESI artefacts are adducts when neutral analytes interact with an ionizing ion other than a proton. The most common adduct in positive ESI is with sodium which forms an M+Na ion. Other common adducts are with ammonia, water and potassium in positive ESI and in negative ESI with formate, acetate and chlorine. The formation of some of these adducts can be limited by thoroughly desalting samples prior to analysis (de Hoffmann and Stroobart, 2007).

1.3.1.2.5.2 Atmospheric Pressure Chemical Ionization

APCI sources are similar to that of the ESI, typically a larger heater is incorporated to enable temperatures to reach 600˚C to enable a more rapid desolvation of mobile phase. Furthermore the gas phase mobile phase is ionized via corona discharge as in CI. It is the interaction between these ions that ionizes the analytes in a similar manner to that of chemical ionization but only at atmospheric pressure. In positive ionization a proton is transferred from the charged mobile phase to the analytes and vice versa in negative mode (de Hoffmann and Stroobart, 2007). However, due to the greater temperature of the APCI source many thermo liable metabolites may be degraded in a metabolomic analysis.

1.3.1.2.5.3 Atmospheric Pressure Photoionsation

The APPI source is similar to that of the APCI source and only the corona discharge is replaced by a photon source, typically a UV lamp (Robb et al., 2000). The interaction between the photons and analyte in gas phase induces ionization either through protonation or deprotonation. APPI is of particular use for analytes not ionized by ESI and APCI, and these include non-polar compounds such as polyaromatics and conjugated compounds (Lien et al., 2009, Robb et al., 2000). As with APCI the increased thermal degradation of analytes may hinder metabolomic analysis, in addition to having a much reduced ionization efficiency relative to APCI and ESI (de Hoffmann and Stroobart, 2007).

1.3.1.2.6 Mass analysers

To determine the mass of analytes a mass analyser is required and it is into this that ions in the gas phase are directed. The property that is determined by the mass analyser is the mass to charge ratio (*m/z*) of the ion rather than the mass. Due to the different ionization techniques and applications utilizing mass spectrometry a number of mass analysers have been developed, each with its pros and cons. The sensitivity and mass resolution of mass analyzers is continually being improved upon. Mass resolution refers to the precision of the mass measurement, low resolution (≤5000) mass spectrometers only provide whole integer (nominal) mass. This makes metabolite identification more difficulty as many metabolites will have the same nominal mass. An example of this would be creatinine ($C_6H_{13}NO_2$) and leucine ($C_4H_9N_3O_2$) both common urinary metabolites and both with a nominal mass of 132 Da. A high resolution mass analyser will provide a more precise mass measurement, for example a mass analyser with >10,000 mass resolution will return mass of 132.0808 and 132.1019 for creatinine and leucine respectively thus allowing a accurate molecular formula to be determined (Watson, 2013). However, sensitivity and resolution are indirectly proportional, and improving one often sacrifices the other so for many MS platforms a compromise must be

reached (de Hoffmann and Stroobart, 2007). Mass analysers can be split into several classes, and single mass analyser are mass spectrometers made up of one mass analyser such as single quadrupole or time of flight. Quadrupole analysers can be added in series in a triple quadrupole or used to produce hybrid mass analysers such as a quadrupole time of flight (QTOF).

1.3.1.2.6.1 Quadrupole

Quadrupole mass analysers compose of 4 parallel hyperbolic rods, of which 2 use radio frequency (*r.f*) rods and 2 use direct current (*d.c*) rods to isolate selected parent ions based upon *m/z* and propel them the detector (Glish and Vachet, 2003, de Hoffmann and Stroobart, 2007). Ions falling outside the desired *m/z* impact the quadrupole rods and fail to reach the detector. However, full spectrum scans are possible by continual variation of the *d.c* property. Quadrupole mass analysers are low resolution with only nominal mass accuracy (de Hoffmann and Stroobart, 2007) making them a poor choice for metabolomics analysis.

1.3.1.2.6.2 Triple Quadrupole

A triple quadrupole is a linear arrangement of 3 quadrupole mass analysers where quadrupoles 1 and 3 selectively filter specific ions while quadrupole 2 is operated with only *r.f* and acts as a collision cell to create fragments. These are then focused and directed towards quadrupole 3 which selectively filters specific daughter ions. As with the single quadrupole instruments mass resolution is poor with only nominal mass accuracy. However, sensitivity is improved and the technique is widely used for quantitative analysis using the parent ion and a fragment for confirmation of identity in a process known as multiple reaction monitoring (MRM) (de Hoffmann and Stroobart, 2007). This analytical platform is particularly useful in targeted metabolomics studies where the metabolites of interest are well established and the effect of metabolic disruption is being quantified.

1.3.1.2.6.3 Time of Flight (TOF)

Time of flight mass spectrometry has been utilized for the best part of 60 years and still remains a widely used mass spectrometry technique. The basis behind the time of flight mass analyser is that an ion's velocity in a vacuum is indirectly proportional to its *m/z* (Guilhaus, 1995). This means that the mass can be determined by measuring the time taken for an ion to travel a known distance. Unlike the quadrupole mass analysers, a strictly defined time period is required meaning ions are pulsed into the mass analyser enabling the time of flight to the detector can be measured (Glish and Vachet, 2003, de Hoffmann and Stroobart, 2007). Linear TOFMS directs ions down a flight tube towards a detector, but this technique has a limited resolution of just 5000 (de Hoffmann and Stroobart, 2007). Increased resolution can

be achieved using reflectors which reflect ions back towards the inlet of the detector. This effectively doubles the drift distance of the ions allowing greater mass resolution and is known as “V-mode” due to the shape of the ion flight path (Glish and Vachet, 2003). Additional reflectors can be added to create “W-mode” which has even greater resolution. However, as with most resolution boosting techniques there is a subsequent drop in sensitivity. The implementation of reflectors in TOFMS instruments increase resolution by up to 5 fold making them high resolution mass spectrometers (de Hoffmann and Stroobart, 2007) meaning ions of similar or same nominal masses can be resolved adding additional information to the sample (Chernushevich et al., 2001), which is ideal from a metabolomics standpoint. Due to their sensitivity and high degree of mass accuracy, TOFMS has risen to the forefront of metabolomics analysis (Theodoridis et al., 2012, Gika et al., 2014).

1.3.1.2.6.4 Quadrupole Time of Flight (QTOF)

The QTOF represents a hybrid mass analyser, where the collision cell of a quadrupole is combined with the high resolution of a time of flight mass analyser (Chernushevich et al., 2001). The development of this hybrid analyser closely followed that of the ESI development, and as such they are a very common and popular combination which allows both qualitative and quantitative analyses to be performed (Chernushevich et al., 2001). When implemented in full scan mode, which is the main interest for metabolomics analysis, the quadrupole operates with only the *r.f* component to focus the ion beam before entering to TOF component. The quadrupole may also be used as a mass analyser as described earlier and adding this application to a TOFMS allows fragments to be analysed at high resolution increasing confidence in ion identification which is an important feature in metabolomics analysis. (Want et al., 2007, Chernushevich et al., 2001, Moco et al., 2007).

1.3.1.2.6.5 Ion trap

Ion trap mass spectrometry was introduced by Paul and Steinwedel in the 1950s. Once ionized, ions become trapped by an initial *r.f.* frequency and a fixed *d.c* voltage. The means all ions within a given *m/z* range become trapped within the mass analyser; the mass range is determined by the *r.f* voltages applied. The movement of ions depends upon the design of the ion trap. Initially ion traps utilised a circular electrode around which the ions travelled and these were known as Paul traps. In a 2D trap, ions oscillate up and down a linear four rod quadrupole each end of which is capped with a reflector to reflect ions back down the quadrupole (March, 1997, de Hoffmann and Stroobart, 2007). Prior to and subsequent to ionization, analytes are subjected to continual collision with helium gas to dampen the energy of these analytes to further increase control and focussing of ions (March, 1997). Ions are

ejected from the ion trap by ramping the *r.f.* amplitude causing ions of specific *m/z* ratios to be ejected leaving the desired ions to be detected (March, 1997, de Hoffmann and Stroobart, 2007). These mass analysers however typically offer good resolution but fail to provide a high degree of mass accuracy compared to TOF instruments due to the number of ions present in the trap known as the space charge effect, in addition to the gas pressure in the ion trap (Schwartz et al., 1991, Zubarev and Makarov, 2013, Gorshkov and Zubarev, 2005). This makes the ion trap better suited to targeted analysis (Zubarev and Makarov, 2013).

1.3.1.2.6.6 Orbitrap

The Orbitrap is the latest in mass analyser technology, invented by Makarov in the last decade of the 20th century. In an Orbitrap instrument, ions enter from an ESI source and pass into a storage quadrupole which works in a similar fashion to a 2D ion trap. In this case all ions are expelled into the analyser rather than a select few as this step is required to convert the continuous ion production of an ESI into pulsed packages of ions required for Orbitrap analysis (Hu et al., 2005). Ions enter the orbitrap via a rapid expulsion from the storage quadrupole, creating a very narrow package of ions, much smaller than in other mass analysers. Ions enter the Orbitrap at a tangent in a gap between a central spindle shaped electrode and an outer electrode. When a voltage is applied between both electrodes, ions are forced towards the central electrode while their velocity imparted by injection from the storage trap causes a centripetal force, forcing the ions around the central spindle electrode (Zubarev and Makarov, 2013). The shape of the electrode creates an axial electrostatic field forcing ions toward the centre of the trap causing them to oscillate at their harmonic frequencies creating an ion current image which is picked up by the outer electrode. This ion current is then amplified and translated from analogue to digital data before pre-processing and data acquisition (Hu et al., 2005, Zubarev and Makarov, 2013, de Hoffmann and Stroobart, 2007). The resolution of Orbitrap mass spectrometers exceeds 250,000 and offers a significant improvement over TOFMS platforms.

1.3.1.2.6.7 Fourier transformer-ion cyclotron resonance

Fourier transformer-ion cyclotron resonance (FT-ICR) mass spectrometry was invented by Comisarow and Marshall in 1974 (Marshall et al., 1998). This mass analyser is the highest resolution to date with resolution exceeding 1,000,000 (Brown et al., 2005). Here ions are trapped in a magnetic trap but are in a constant circular motion. Ions are excited to their cyclotron resonance frequency by an orthogonal oscillating electric field (Marshall, 1985, Amster, 1996). These cyclotron resonance frequencies are unique for each *m/z* and this resonance is detected by ion image detectors which are discussed later in this chapter. Due to

the fact that the oscillating frequency is a function of the *m/z* and size of the magnetic field applied to the ion, mass resolution is increased by increasing the strength magnetic field, and this is typically achieved using superconducting magnets (Brown et al., 2005, Marshall et al., 1998, Amster, 1996). However to reduce the number of collisions within the mass analyser which may have a detrimental effect of resolution a high vacuum is required and a limited number of ions in the analyser (de Hoffmann and Stroobart, 2007). FT-ICR has potential in metabolomics due to its very high mass resolution and sensitivity, which is increased as the ions can be measured over an extended period of time. The very high resolution means many masses of the same nominal mass can be resolved, thus reducing the requirement for chromatographic separation to some degree (Brown et al., 2005). This makes FT-ICR ideal for high through-put analysis using direct infusion analysis which will be further discussed in this chapter.

1.3.1.2.7 Detectors

Upon leaving the mass analyser, ions need to be detected in order for mass spectra data to be collected. The most commonly used detectors are electron multipliers; here ions hit the detector and trigger an electron cascade amplifying the original ion impact. These can be further focused and amplified by directing the electron cascade down a channel, triggering further electrons to be emitted upon impact with the channel walls. Multichannel plates are commonly utilized in TOF and quadrupole; here channels of 10 µm in diameter and several millimetres in length are present on the surface of a plate or disk. Once an ion strikes the entrance a cascade of electrons are amplified down the channel and directed onto an anode for detection of the charge. However there is a significant area on which an ion can strike but not produce an electron cascade thus limiting the sensitivity of the detector (Koppenaal et al., 2005). In addition, these detectors detect only one ion strike at a time per channel, again reducing overall sensitivity and causing saturation of the detector. The mass spectrum is typically generated using a time-to-digital converter which relates the number of ions detected to the intensity of the peak response (de Hoffmann and Stroobart, 2007).

Trapping instruments typically utilise an image current detector. Despite being trapped, ions in orbitraps and ion traps are in constant motion. Within a magnetic field the orbit radius and frequency is a property of their *m/z*. This causes each ion to emit an *r.f.* signal that is unique to the *m/z* of the ion. These measurements are typically very accurate and account for the high mass resolution of these mass analysers (Koppenaal et al., 2005).

1.3.1.3 Direct infusion mass spectrometry (DIMS)

Direct infusion of sample into a mass spectrometer is a very rapid method for metabolomics, making it possible to screen 100-1000s of samples per day (Scalbert et al., 2009, Fuhrer and Zamboni, 2015). A recent development that enables very high resolution and increased metabolome coverage, whilst providing accurate mass measurement is selected ion monitoring (SIM) stitching. This technique infuses a sample into a high resolution orbitrap or fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS) where a narrow mass window of 30 m/z are released into the mass analyser from a storage quadrupole. A series of overlapping SIM windows are subsequently stitched together and has been shown to increase the number of metabolites detected while maintaining a high degree of mass accuracy (Southam et al., 2007). However, even with SIM stitching DIMS suffers from ion suppression due to a lack of chromatographic separation. Ion suppression is a common problem in mass spectrometry; this is where a number of analytes enter the source at the same time. This is either a result of direct infusion or chromatographic columns eluting multiple peaks at the same time; in the later case this is termed co-elution. All analytes entering the source compete for ionisation, thus ionisation efficiency for each analyte is reduced meaning fewer of their ions are generated and subsequently enter the mass analyser to be detected (Annesley, 2003). This means that low abundance metabolites co-eluting with high abundance metabolites are not detectable, thus limiting the number of metabolic pathways analysed.

1.3.2 Hyphenated techniques

Due to the inherent problems with DIMS caused by a lack of separation many metabolomic studies utilize chromatographic separation prior to MS analysis. As with the detection method the chromatographic method chosen influences the sensitivity and the class of compounds detected (Issaq et al., 2009). To achieve the benefit of chromatographic separation mass spectrometers are linked to the chromatographic instrument to create a hyphenated platform. The two most commonly utilized in metabolic studies are GC-MS and LC-MS (Issaq et al., 2009, Lenz and Wilson, 2007, Zhang et al., 2012a).

1.3.2.1 Gas Chromatography-Mass Spectrometry

Gas chromatography (GC) is an analytical technique used to separate vaporised compounds without thermal degradation of target compounds. Compounds are separated using an inert carrier gas such as helium as the mobile phase which flows through a stationary phase bonded to a GC column. A typical GC column is a reverse stationary phase allowing non polar compounds to elute before polar compounds. The rate of the elution is dependent upon

the temperature gradient of the GC, the greater the temperature the greater the rate of elution. GC-MS has proved to be very popular in the field of metabolomics with many studies utilizing the technique. This hyphenated technique utilizes chromatographic separation which is a vital component of mass spectrometry based metabolomics (Issaq et al., 2008a). GC-MS has been proved to be a highly reliable and reproducible technique for metabolomic analysis (Issaq et al., 2008a). The ability to selectively fragment ions and the ability to search mass spectral libraries make metabolite identification less arduous and much quicker (Scalbert et al., 2009, Dettmer et al., 2007). The use of GC-TOFMS opposed to more traditional GC-MS increases mass accuracy and therefore confidence in metabolite identification and has been used to good effect in a number of metabolomic analyses (Liu et al., 2010, Ma et al., 2011).

A number of recent advancements have significantly improved GC-MS analysis for metabolomics including 2 dimensional GC separation and linkage to high resolution TOFMS platforms. In 2D GC separation, metabolites eluted from one column are trapped and released via thermal modulation onto a second, typically more polar, column. This has a combined effect of increased chromatographic separation, peak capacity and improved signal to noise ratios (Hagan et al., 2007).

For some studies, several features of GC-MS can make it preferential to NMR and DIMS analysis as the separation using the GC increases sensitivity and resolution of the platform allowing for a broad range of metabolites to be analysed (Scalbert et al., 2009). However, whilst GC separation is suitable for a wide range of compounds it is limited to those that are volatile or can be made volatile by derivatization (Ramautar et al., 2009, Kaddurah-Daouk et al., 2008, Dettmer et al., 2007) and to thermo stable compounds due to the high temperatures utilised during chromatographic separation (Ramautar et al., 2009). These properties limit the analysis to a smaller selection of the metabolome than some of the other techniques available.

1.3.2.2 Liquid Chromatography-Mass Spectrometry

In recent years LC-MS has risen to the forefront of metabolomics analysis, capable of separating a dynamic range of metabolites and linked to high resolution mass spectrometers (HRMS) (Dettmer et al., 2007, Gika et al., 2014, Moco et al., 2007). LC-MS offers many advantages over NMR, and DIMS including its increased sensitivity, separation and resolution (Dettmer et al., 2007). In comparison with GC-MS, separation and detection are independent of sample volatility meaning sample derivatization is rarely required, although LC-MS typically struggles with non ionizable compounds such as hydrocarbons (Werner et al., 2008). Liquid

chromatographic separation of the metabolome allows distinctions between polar, mid polar and nonpolar metabolites to be made and reduces ion suppression within an ESI source.

Metabolomic analysis utilizing LC systems initially utilised high performance liquid chromatography (HPLC) platforms (Issaq et al., 2008a, Wilson et al., 2005). These provide separation of a wide range of metabolites, HPLC however is limited by factors including: long run times, limited back pressure (2000-4000psi) and particle sizes greater than 2µm (Swartz, 2005, Churchwell et al., 2005). All of these factors impact upon the resolution and peak capacity of the chromatographic separation meaning that metabolites are more likely to co-elute and ion-suppress each other (Gika et al., 2014). To counter this, new column technology has been developed using columns with a sub 2µm particle size and reduced internal column diameters, and the combined effect of these leads to high back pressures of around 8000-12000psi (Issaq et al., 2008a). This results in increased resolution, peak capacity and reduced run times (Wilson et al., 2005, Ryan et al., 2011, Swartz, 2005, Nguyen et al., 2006). The incorporation of these columns and pumps capable of handling the high back pressures led to the development of UHPLC platforms (Sterz et al., 2012, Wilson et al., 2005, Swartz, 2005, Churchwell et al., 2005).

To achieve adequate separation and retention of the urinary metabolome several column chemistries need to be considered as no single column chemistry is capable of encompassing the entire urinary metabolome (Zhang et al., 2012b, Spagou et al., 2011, Gika et al., 2008a). As such, the column chemistry that allows for the detection of the greatest number of metabolites or a combination of methods to give the broadest range of urinary metabolites is required (Theodoridis et al., 2008). To date reverse phase (RP) column chemistries dominate metabolomic analysis (Issaq et al., 2008a, Idborg et al., 2005, Gika et al., 2014) with hydrophilic interaction liquid chromatography (HILIC) being reserved for more directed metabolomics of polar compounds (Gika et al., 2014).

The stationary phase of RP columns is non polar, typically utilising trifunctionally-bonded C18 alkyl chains bound to a silica or hybrid hydrocarbon-silica particle (Neue et al., 2006). In RPLC the polarity of the mobile phase decreases over the LC gradient (Issaq et al., 2008a). This enables analytes to separate based upon hydrophobicity as well as interaction with any free silanol groups on the column phase. Due to the non polar nature of the stationary phase, polar analytes are poorly retained and elute in the solvent front, these metabolites include amino acids, nucleotides and small organic acids to name a few (Gika et al., 2008a, Zhang et al., 2012b, Cubbon et al., 2007). These polar metabolites are then followed

by less polar metabolites such as conjugated pharmaceuticals and steroids, then none polar compounds such as unconjugated steroids, phospholipids and fatty acids.

In order to counter the problems arising from lack of retention of highly polar ionic analytes on RPLC, HILIC stationary phases have been used (Cubbon et al., 2007, Gika et al., 2008a, Spagou et al., 2011). The mechanism by which HILIC separates these compounds is the opposite of the more conventional reversed phase chromatography and is sometimes incorrectly referred to as normal phase chromatography (Yoshida, 2004). While there are other well established methods for separation of polar compounds such as ion exchange chromatography (IEC), the mobile phases used for HILIC analysis are more compatible with electrospray ionization mass spectrometry (ESI-MS) (Boersema et al., 2008). Under HILIC conditions the unbound stationary phase particle is polar and the mobile phase increases in polarity over the course of the LC gradient (Issaq et al., 2008a). This allows the polar compounds to elute with the increasing polarity of the mobile phase, while non polar compounds typically elute early on in the run. The theory behind the separation in HILIC columns is still debated, however there are two commonly agreed upon methods which are thought to complement each other. It has been observed that there is a layer of immobile water coating the stationary phase of the HILIC column (Hemstrom and Irgum, 2006), this leads to the theory that liquid-liquid extraction of compounds occurs between this water layer and the more hydrophobic mobile phase. The second theory is that hydrogen bonding between compounds and the water layer allows for this separation. The reality is that it is likely a combination of the two theories that explain the mechanism behind HILIC retention (Chauve et al., 2010, Novakova and Vlckova, 2009).

For the analysis of urine a number of studies have been carried out to compare the chromatographic properties of both RPLC and HILIC in metabolomic analysis (Cubbon et al., 2007, Gika et al., 2008a, Spagou et al., 2011, Zhang et al., 2012b). Gika et al., (2008a) and Zhang et al., (2012b) demonstrated using principal component analysis (PCA) that there are qualitative differences in metabolome coverage between RP and HILIC analysis. While HILIC analysis has been shown to improve retention and subsequent detection of polar metabolites, the non-polar metabolites suffer from lack of retention and reduced sensitivity. As such, reversed phase and HILIC columns are considered complementary and together provide greater metabolome coverage than either on its own (Want et al., 2010, Cubbon et al., 2007, Spagou et al., 2011).

1.3.2.2.1 nanoflow LC-nanospray ESI-TOFMS

The latest developments LC-MS have led to the miniaturisation of conventional LC-MS and the development of nanoLC-nanoESI-MS (nLC-nESI-MS). As the name suggests flow rates here are on the nL/min scale and has been defined as LC platforms that deliver flow rates of 10-1000 nL/min with ESI emitters of between 10 and 150 µm (Chervet et al., 1996). These nano scale platforms result in a significant improvement in terms of sensitivity and offer great potential to metabolomic researchers looking to encompass as larger a proportion of the metabolome as possible. In addition, the very small injection volumes used allow analysis of precious and limited samples such as tissue biopsies (Jones et al., 2014a). The nano flow rates mean much lower volumes of solvents are required, up to 1000 times less than conventional LC (Hernandez-Borges et al., 2007), thus reducing the cost of analysis in addition to reducing the environmental impact of metabolomic analysis (Hernandez-Borges et al., 2007). For the past several years there has been a call for nLC-nESI-TOFMS to be implemented in metabolomics as a method of increasing metabolome coverage particularly of low abundance metabolites (Griffiths et al., 2007, Gika et al., 2014, Pendyala et al., 2007, Want et al., 2006, Want et al., 2005, Want et al., 2007, Ek et al., 2010). However, very few studies have followed up on this. Conversely in the sister -omics field of proteomics, nLC-nESI-MS has risen to the forefront of protein analyses (Fischer et al., 2013).

The source of the increased sensitivity of nLC-nESI-MS platforms is two-fold compared to conventional platforms. The initial increase in sensitivity in systems utilizing chromatography comes from the increased chromatographic separation, thus reducing co-elution of compounds and subsequent ion suppression (Schmidt et al., 2003). In addition, due to the much smaller flow rate the amount of chromatographic dilution is significantly decreased thus allowing metabolites to elute at a greater concentration, leading to improved peak intensity (Gama et al., 2013, Fanali et al., 2007). However the main source of the increased sensitivity is in the formation of the nano spray ESI. As a result of the low flow rate from the nanoflow LC system, the ionization efficiency of ESI is significantly improved. Using nanoESI emitters with and internal diameter (id)of 10-150 µm, the spherical droplets formed in the ESI plume are 100-1000 fold smaller than the typical 1-2 µm diameter droplets of a conventional ESI emitter (Wilm and Mann, 1996). The process of ESI is an evaporative one, due to the significantly smaller plume droplets the volume of emitted droplets is significantly decreased relative to conventional ESI. This significantly increases the rate of desolvation of said droplets, and due to this ions are partitioned into the gas phase at a much greater rate and thus more enter the mass analyser (Marginean et al., 2008, Marginean et al., 2014, Gangl

et al., 2001, Karas et al., 2000). The relationship between initial droplet volume and efficiency has been extensively described elsewhere (Juraschek et al., 1999, Wickremsinhe et al., 2006). The result of this increased ionization efficiency means that up to 500 times more ions become ionized in a nESI source relative to conventional ESI (Wilm and Mann, 1996).

Each of the 3 major analytical platforms, NMR, LC-MS and GC-MS, used for metabolomic research has its advantages and disadvantages, and it is generally accepted they complement each other (Wilson et al., 2005) . The use of LC-with high resolution ESI-MS is widely utilised in the field of metabolomics however much remains to be done to improve current metabolomic analysis, particularly in the case of low abundance urine metabolites.

1.3.3 Current analytical limitations in urinary metabolomics

There are significant challenges to overcome to achieve a sensitive and reliable urine metabolomics methodology. These include the detection of low abundance metabolites (German et al., 2005, Dettmer et al., 2007), such as eicosanoids and estrogens which are important signalling molecules in the body and associated with several disease states when deregulated. To date these concerns have been muted in favour of speed of analysis, with the aim of being able to run more samples in any given time therefore allowing more metabolomic studies to be done (Gika et al., 2014). This however has a drawback in that only the most abundant metabolites present are analysed, potentially leaving low abundance metabolite variation to go undetected, and thus reducing coverage of a potentially important part of the urinary metabolome.

1.4 Analysis of analytical datasets in metabolomics studies

The -omic sciences typically generates vast data sets with a large number of variables present within samples and the different sample groups. Multivariate analysis (MVA) is used to assess similarities and/or differences between metabolic profiles of different groups and indicate which metabolites are driving the discrimination (Issaq et al., 2009). Multivariate analysis allows a visual representation of the data, to quickly assess the relationships between sample groups. In addition, it allows outliers to be identified quickly and is routinely used as a way of ensuring quality control of the acquired data set. Data analysis and metabolite identification account for the final steps of a global metabolomic work flow (Figure 1.2).

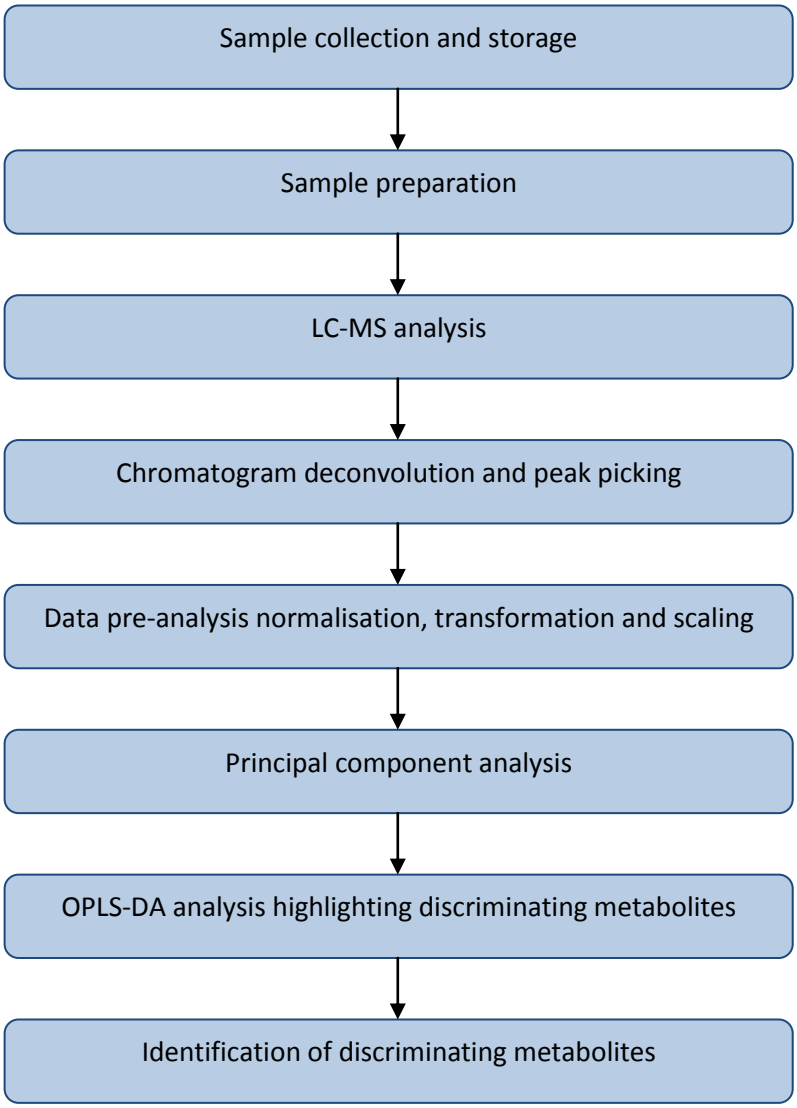


Figure 1.2: Typical work flow of a LC-MS based metabolomic study

1.4.1 Pre-processing of data sets

Prior to multivariate analysis, MS data sets require a pre-processing step known as binning which provides retention time (Rt), *m/z* and ion intensity in a process known as peak picking and chromatogram deconvolution. The retention time and *m/z* are typically combined as the primary identity of the variable, while the ion intensity is the measure of ion concentration. This pre-processing reduces the complexity of the data from tens or hundreds of thousands of ions to a few thousand per batch; each *m/z* x Rt is picked using criteria selected by the user. These criteria include the amount of noise reduction, number of counts required to be classed as a peak, the retention time window and the mass window (Liland, 2011). These data are subsequently organised to give each metabolite and its normalised intensity (termed the variables) in columns and individual samples in rows (termed the

observations) (Wang et al., 2012b). Due to the large dynamic range of the metabolome each metabolite may be present at concentrations orders of magnitude different from each other. This means that abundant metabolites with high variance will skew the analysis of low abundance metabolites. To counter this, metabolomic data sets are typically transformed, centred and scaled (Liland, 2011, Eriksson et al., 2006). The use of centering centres all variance on zero as opposed to the mean intensity. Thus high and low abundance metabolites are treated equally with the intensity variation being the focus of the statistical analysis (van den Berg et al., 2006). The most common form of scaling is a log transformation which, in addition to reducing the impact of very high abundance compounds, typically results in skewed data becoming more normally distributed (Liland, 2011, van den Berg et al., 2006). However by reducing the impact of high abundance compounds, the baseline noise level is increased making the efficiency of peak picking process critical (Liland, 2011). In metabolomic data sets, scaling is also used to reduce the effect of varying fold changes between different metabolites. This is achieved by dividing the variable by a scaling factor and this results in an increased impact of variables with small variation, which typically would go unnoticed as a result of a few metabolites with a large degree of variation between sample groups. For metabolomic analysis pareto scaling is typically used whereby the mean is subtracted and the square root of the standard deviation is used as the scaling factor, and this reduces large fold changes more than small fold changes. This means large fold changes are less dominant in the statistical analysis (van den Berg et al., 2006, Eriksson et al., 2006, Gika et al., 2008b).

1.4.2 Unsupervised principal component analysis

Typically the first statistical analysis performed is a principal component analysis (PCA); this is an unsupervised approach. The PCA offers an insight into how the observations are related to one another (Eriksson et al., 2006, Liland, 2011).

A PCA is constructed in K-dimensional space whereby there are as many dimensions as there are variables, with each variable being a 1 co-ordinate axis. Following this each observation is plotted in K-dimensional variable space leading to a “swarm” of points. This subsequently allows for the calculation of the first principal component which is a straight line passing through the average point in K-dimensional space, as such it represents the maximal variation in the direction of the data. A further component, the second principal component, is added again passing through the average point orthogonal to the first component and accounting for much of the residual variation. The combination of the first and second component represents a hyperplane in K-dimensional space, the lowest dimensional representation of the multivariate data. The co-ordinate value for each observation is known

as “score”, which lends its name to the visualisation of the 1st and 2nd component; the scores plot. The scores plot allows for the observation of variable grouping/clustering (Figure 1.3), and to further explore this grouping a loading plot can be constructed which visualises the influence the loading variable has on the scores plot (Eriksson et al., 2006). There are usually many more than two components that can be fitted to the data set and these should be investigated but often are a result of noise and do not offer reliable explanations of variability within a dataset (Liland, 2011).

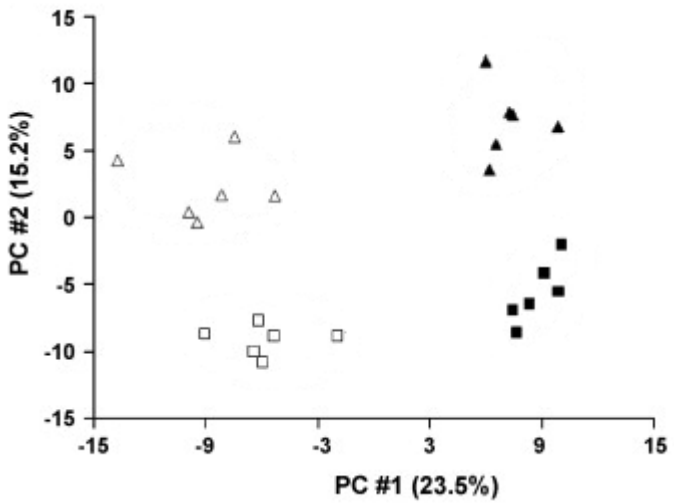


Figure 1.3: A typical PCA scores plot used to understand how sample groups relate to one another

PCA scores plot of male (squares) and female (triangles) rats on high (solid fill) and low (open) dose drug regimes. Scores plot indicates significant urinary metabolomic differences between the four samples groups. Figures in brackets indicate the amount of variation explained by the model for each component. Adapted from (Normalisation strategies for metabonomic analysis of urine samples, 2009) (Warrack et al., 2009)

1.4.3 Supervised partial least square discriminant analysis (PLS-DA) and orthogonal partial least square discriminant analysis (OPLS-DA)

PCA analysis determines the directionality of the major sources variation in the dataset. However, this may be unrelated to the maximum variation between samples groups. Supervised analysis using PLS-DA uses knowledge of sample groups to maximise covariance explained by latent variables associated with sample group (Eriksson et al., 2006, Liland, 2011). This, for example, would maximise separation between different dose groups or disease progression, however the first latent variable may not explain the greatest source of variation between dose or disease progression groups (Wiklund et al., 2008). To counter this, pair wise OPLSA-DA rotates the dataset to ensure the maximal variation between groups now lie on the 1st predictive component, and any variation not correlated to groups are found in orthogonal components (Wiklund et al., 2008). OPLS-DA can be visualised by constructing an S-scatter plot, and this visualises the influences individual variables have on the model. S-plots combine

both the covariance and the correlation of each variable on the model, that is to say the magnitude and reliability of the effect (Wiklund et al., 2008, Westerhuis et al., 2010). The use of S-plot allows individual metabolites driving separation between groups to be highlighted for identification and quantification (Figure 1.4). These are highlighted due to their high covariance and high correlation, however, those with a high correlation and low covariance may still be of interest though should be thoroughly examined to ensure these are not artefacts due to noise and analytical variability (Wiklund et al., 2008).

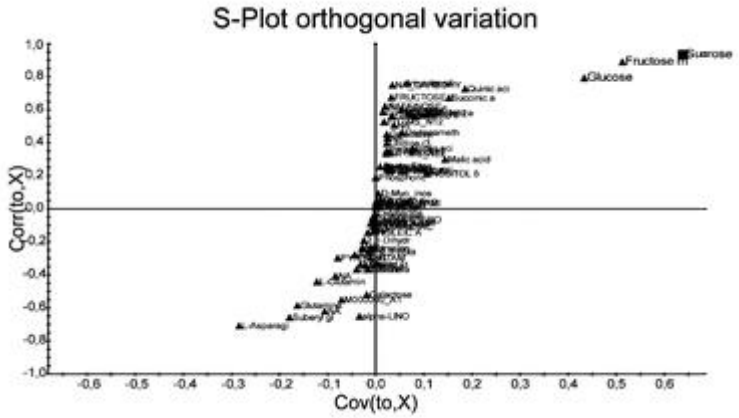


Figure 1.4: Example of S-plot analysis of a transgenic poplar line relative to wild type poplar
Different sugars are highlighted as the variables driving separation between the two plant lines. Cov= covariance. Corr= correlation. Adapted from (Visualization of GC/TOF-MS-Based Metabolomics Data for Identification of Biochemically Interesting Compounds Using OPLS Class Models, 2008) (Wiklund et al., 2008)

The structure of the discriminating metabolites are identified using an array of techniques (Figure 1.5) and initial accurate mass measurements allow for an elemental composition to be calculated. These data can then be used to search mass spectral databases such as the human metabolome database (HMBD) (Wishart et al., 2013), mycompoundID (Li et al., 2013a) and Metlin (Tautenhahn et al., 2012a). Many of these databases contain mass fragment patterns which can be compared to the unknown metabolite peak. In addition authentic standards can be analysed as a final check to ensure retention times match and also for generation of fragmentation pattern should they not be available in databases (Moco et al., 2007, Dunn et al., 2013).

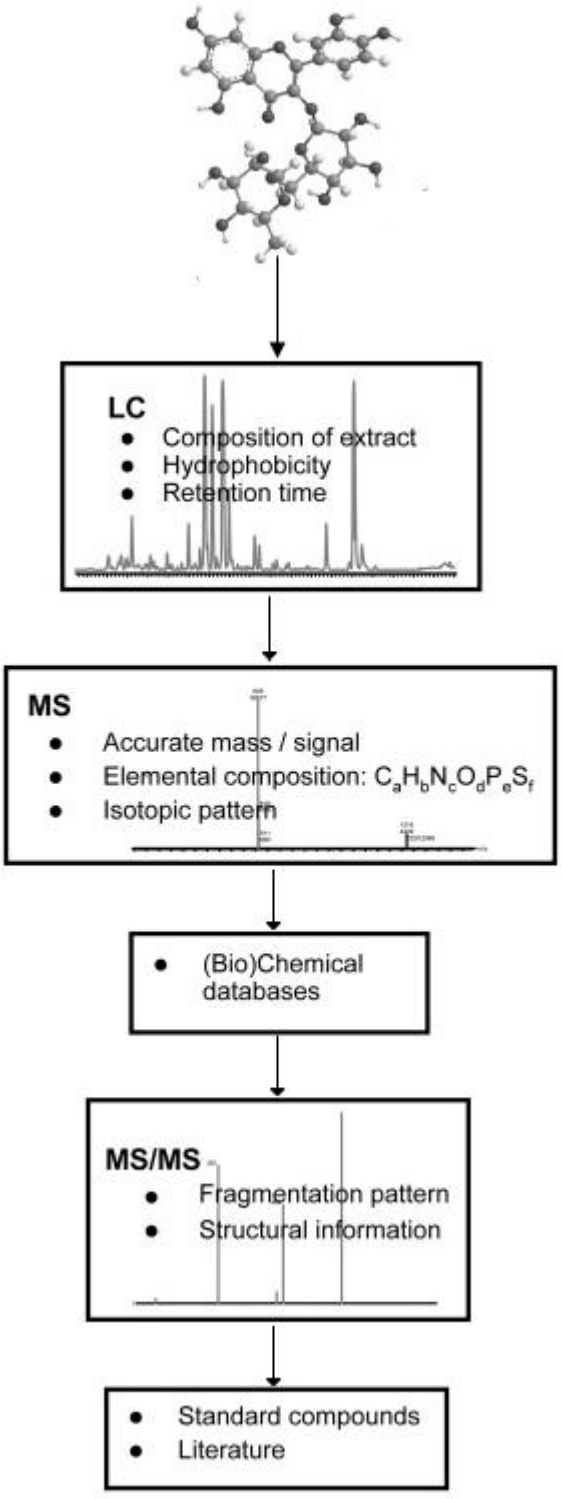


Figure 1.5: Schematic for the identification of unknown metabolites

Adapted from (Metabolomics technologies and metabolite identification, 2007) (Moco et al., 2007)

1.5 Urinary metabolomics methodology

In studies of human health and disease, the use of non-invasive techniques for biomarker discovery and subsequent monitoring in patients are desirable to increase patient participation, while providing a plentiful supply of urine sample (Want et al., 2010, Ryan et al., 2011). Urine is already commonly used in clinical laboratories (Delanghe and Speeckaert, 2014) meaning collection of urine for metabolomic analysis is of little extra effort for clinicians. The urinary metabolome could potentially be used to investigate metabolic consequence of disease for the entire body due to it being a major excretory route of water soluble metabolites, meaning metabolites from many biochemical pathways are present in urine after their formation around the body (Want et al., 2010). Urine is produced in the kidneys and requires three major processes; these are filtration, reabsorption and secretion. As blood enters the kidney via the afferent arteriole it is filtered by the glomerulus. Here small molecules, ions and water pass through the glomerulus, whereas larger blood components such as proteins and cells, which are too large to pass through the glomerulus, leave the kidney via the efferent arteriole. The glomerular filtrate passes through the proximal convoluted tubule where selective reabsorption process occurs. In this region of the kidney most of the water is reabsorbed via osmosis. In addition, most mineral salts, glucose, amino acids and other metabolites are reabsorbed either by diffusion or active transport into the capillaries surrounding the kidney. Further reabsorption of water occurs as the filtrate passes through the loop of Henle, prior to passing into the distal convoluted tubule. A similar reabsorption procedure occurs here as described in the proximal convoluted tubule only the level of water, ions and metabolites reabsorbed is greatly reduced as most of the vital components have already been reabsorbed. The secretory process occurs in both convoluted tubules, it is by this route that waste products such as creatinine, xenobiotics, urea, hydrogen ions and bile acids eliminated into the urine. These molecules pass from the blood across the tubule membrane, primarily via active transport, into the glomerular filtrate. The remaining filtrate which now contains the bodies excess water, mineral salts, metabolites and waste products passes into the collecting duct. This is subsequently stored in the bladder as urine prior to elimination (Pocock and Richards, 2006). The more lipophilic metabolites such as lipids and phospholipids are excreted mainly via the faeces. In addition to the endogenously derived metabolome, xenobiotics consumed in day to day life such as dietary compounds, pharmaceuticals and environmental pollutants are also detected in urine. The presence of both endogenous and exogenous metabolites means that urine metabolomics may be implemented for disease biomarker discovery (Ganti et al., 2012a, Ganti et al., 2012b, Cheng et al., 2012, Ganti and Weiss, 2011, Danielsson et al., 2011, Kim et al., 2011), drug discovery and

characterisation (Satheeshkumar et al., 2012, Kaddurah-Daouk et al., 2008, Lindon et al., 2007), determination of nutritional status (Xie et al., 2013, Goodacre, 2007, Andersen et al., 2013, Llorach et al., 2012, O'Sullivan et al., 2011, Kussmann et al., 2006) and effects of environmental toxicants (Miller, 2007, Morrens et al., 2012). To date over 3000 metabolites have been characterised in human urine, and this is considered the minimum number of metabolites present, as many low abundance metabolites may yet be discovered (Bouatra et al., 2013). Of these 3000, at least 2000 are thought to be exogenous in origin from the diet, pharmaceutical intake or through environmental exposures (Bouatra et al., 2013). Given that the food metabolome encompasses >25,000 metabolites, it suggests that 3000 urinary metabolites is a very conservative estimate (Scalbert et al., 2014).

1.5.1 Storage and stability of urine samples

Analysis of urine rarely occurs immediately following collection due to the large number of samples collected or the time required to prepare them for analysis. This means that suitable sample storage is required to ensure sample degradation does not confound metabolomic analysis (Gika et al., 2008b, Alvarez-Sanchez et al., 2010a). Several studies have been performed to quantify the effect of different sample storage solutions (Gika et al., 2008b, Pasikanti et al., 2008, Saude and Sykes, 2007, Gika et al., 2007). These studies conclude that while long term stability is acceptable at -20°C, storage at -80°C is preferred as these studies were limited to 6 months and longer term stability is unknown (Saude and Sykes, 2007, Gika et al., 2007, Gika et al., 2008b, Want et al., 2010). Storage at room temperature is described by Saude et al., as inadequate for metabolomic analysis, with significant variation seen in the stability of a number of metabolites associated with glycolysis (Saude and Sykes, 2007). Analytical run time for metabolomics analysis can be from several hours to several days during which urine samples should be kept in a refrigerated autosampler at 4°C. Under these conditions urine samples deteriorate but at a slower rate than at higher temperatures (Saude et al., 2007). Studies indicate that it takes up to 48 hours at 4°C before significant alterations in the urinary metabolome can be resolved using PCA analysis and so it is recommended that samples be stored for no longer than 48 hours at this temperature (Gika et al., 2008b, Gika et al., 2007). In addition, it is clear that a number of freeze thaw cycles are involved with metabolomics analysis; samples are usually frozen upon collection, defrosted for sample preparation, refrozen before analysis and again defrosted when the time for analysis arrives. The impact of such freeze thaw cycles was also investigated by Gika et al., (Gika et al., 2008b, Gika et al., 2007), and they reported that no appreciable impact on the urinary metabolome

was observed following 9 freeze thaw cycles. However, it is prudent to limit the number of freeze thaw cycles to as few as reasonable possible (Scalbert et al., 2009).

A further complication is bacterial contamination and metabolism of urinary metabolites. While the urine itself is sterile in a healthy individual, bacterial contamination is introduced from the urethra during urination. It has been shown that freezing samples at -80°C is enough to stop the action of bacterial degradation (Scalbert et al., 2009). The effects of sodium azide as a bacteriostatic compound for use in urine metabolomics has been investigated to help reduce metabolic variation (Snyder and Lichstein, 1940). The use of other bacteriostatic/bactericidal compounds such as sodium fluoride also shows promise (Scalbert et al., 2009).

1.5.2 Sample preparation

Sample preparation for urinary metabolomics is an underappreciated aspect of the experimental work yet plays a vital role in the quality, reliability and coverage of the metabolomic data set (Alvarez-Sanchez et al., 2010a, Gika et al., 2014, Theodoridis et al., 2012). Depending upon the aim and scope of the metabolomic study, different sample preparation methods can be implemented. In targeted approaches very selective sample preparation can be utilized to increase the sensitivity of the analysis to selected compounds (Alvarez-Sanchez et al., 2010b, Fernández-Peralbo and Luque de Castro, 2012). However, in global urine metabolomics a non selective sample preparation is required and currently is limited to centrifugation to remove solid material or dilution followed by centrifugation (Want et al., 2010, Gika et al., 2014). Neat and diluted urine preparations both have merit for metabolomic analysis; neat is unmodified and therefore contains the whole metabolome. However, when neat urine is analysed by UHPLC-ESI-TOFMS, the sensitivity can be adversely affected due to co-elution of peaks and subsequent ion suppression, the high salt content also aids the formation of adducts within the ESI source in addition to fouling the LC column and the ESI source (Waybright et al., 2006, Alvarez-Sanchez et al., 2010b). By diluting urine samples prior to analysis it is possible to reduce this ion-suppression and potentially uncover some lower abundance metabolites which were previously hidden by a high abundance co-eluting metabolite or urinary salts. However, there is also the possibility that low abundance peaks are diluted to below the level of detection of the mass spectrometer. Indeed, this has been observed to be the case in human urine metabolomics analysis (Issaq et al., 2008b, Waybright et al., 2006) but conversely for rodent urine which is proteinatious in nature dilution has been observed to improve sensitivity (Waybright et al., 2006).

Although not yet explored for non targeted urine metabolomic analysis, solid phase extraction (SPE) has potential to improve upon current sample preparation methodologies. Currently SPE is reserved for either targeted approaches or as a basic sample clean up (Fernández-Peralbo and Luque de Castro, 2012). The use of SPE would greatly improve the removal of both ion suppressing salts and proteins, in addition to providing the opportunity to concentrate urine samples prior to analysis (Alvarez-Sanchez et al., 2010b, Fernández-Peralbo and Luque de Castro, 2012). This would enhance detection of low abundance metabolites which would not normally be detected using a neat or diluted sample, in addition to increasing column lifespan and reducing the fouling of the MS source (Fernández-Peralbo and Luque de Castro, 2012).

1.5.3 Urine metabolome analysis reliability and quality control

Metabolomic analysis requires a stable analytical platform to enable reliable results to be generated without the introduction of bias introduced by the LC-MS system. During an analytical run, bias can be introduced due to source fouling causing a drop in sensitivity, meaning as analysis progresses analytical sensitivity decreases. In addition, changes in column performance throughout metabolomic analysis may cause retention time drift which confounds the peak picking process. This can be reduced by careful sample clean up to remove salts and proteins and by analysing samples in a randomised order to spread the effect of any sensitivity loss across all samples groups (Benton et al., 2012).

The final stage of any analytical method development prior to metabolomic analysis is the assessment of analytical reproducibility. This needs to be tested within day, to ensure analysis remains reliable during batch analysis over the course of a day, and between day reproducibility to ensure the same results can be obtained on different days (Gika et al., 2007). For large scale, multisite metabolomic analysis it also becomes important to assess the intra-laboratory reproducibility to ensure all laboratories are providing the same analytical output (Benton et al., 2012).

There are several methods for assessing method reproducibility currently in use; these include single metabolite/standard or whole metabolome variation using quality controls. In addition, the use of quality control samples during a metabolomic batch analysis is widely adopted to assess within batch reproducibility.

A number of studies implement single or multiple internal standards or metabolites to assess retention time and signal intensity reproducibility. This is achieved by spiking urine samples with known standards and analysing them as analytical replicates. The mean retention

times and signal intensities are calculated and reported as a % coefficient of variation (%CV) (Benton et al., 2012). Reproducibility for single ion intensities is generally seen as acceptable when %CV for mean peak area in analytical replicates fall below 20% (Benton et al., 2012, Spagou et al., 2011), whereas for retention time a %CV of less than 2% is desirable (Theodoridis et al., 2012).

Another more commonly used measure of reproducibility is to calculate mean peak areas and %CV for all peaks present in a set of quality control (QC) replicates (Spagou et al., 2011, Want et al., 2010, Gika et al., 2007). These QCs are typically an aliquot from all samples combined to produce a pooled sample. The QC is then injected at the beginning and end of a batch and at random intervals during the analysis of the batch (Theodoridis et al., 2008). In these cases, only peaks present in 80% of the QCs "the 80% rule" are utilised, and of these 70% ideally should have a %CV for mean peak area of <30% (Gika et al., 2007, Want et al., 2010). In addition, PCA analysis of the clustering of QC samples on a scores plot also enables a quick visual representation of analytical variability associated with just analytical variance opposed to biological and sample preparation variation.

A further technique to reduce analytical variability in samples batches run is to allow the ESI source to condition before the analysis of samples. Typically this is done starting each batch with several QC injections, the first few of which will group away from the main QC group on a PCA scores plot but slowly track towards the group over the space of several injections (Want et al., 2010, Gika et al., 2007).

1.5.4 Normalisation

In contrast to other bodily fluids the volume of urine present in each void is not physiologically controlled and instead it is dependent upon the individual's hydration status i.e. water consumption (Warrack et al., 2009, Chen et al., 2013b, Veselkov et al., 2011a). The urine volume is known to vary by up to 15 fold in a healthy individual and this subsequently leads to substantial variation in metabolite concentration (Chen et al., 2013b, Veselkov et al., 2011a). Normalisation is a technique to reduce this biological variation and return representative values for metabolite concentration independent of urine volume (Veselkov et al., 2011a, Chen et al., 2013b, Warrack et al., 2009, Heavner et al., 2006). At present there is no agreed accepted method to normalise for urine concentration and a number of techniques exist but all have their advantages and their drawbacks. The most commonly cited methods in the literature are to normalise to the creatinine concentration, the osmolality, the total mass spectrum signal or total usable mass spectrum signal.

1.5.4.1 Creatinine normalisation

The use of creatinine as a normalisation factor has been widely used as a normalisation method for either targeted or metabolomic analyses of urine. Creatinine is an easily measured either enzymatically or though the colorimetric Jaffe reaction. Creatinine is present in human urine produced from the breakdown of creatine phosphate in muscle tissue (Warrack et al., 2009, Heavner et al., 2006). It does however make the assumption that creatinine excretion is homogenous both between individuals and between samples from a single individual. This however is not the case and a combination of factors such as gender, diet, and physical activity can result in up to a 5 fold variation in creatinine concentration, (Warrack et al., 2009). In addition the health of the individual also plays a role in the urinary creatinine concentration (Warrack et al., 2009, Heavner et al., 2006), with increased creatinine being used as a hallmark of kidney damage (Levey et al., 2003). Due to its common use in metabolomics and clinical laboratories to determine the concentration of urine, urinary metabolite concentrations are typically presented relative to the molar concentration of urinary creatinine. An example of this is hippuric acid, a common urinary metabolite, has been reported to be present at 298.5±276.8 μM of hippuric acid per mM of creatinine (298.5±276.8 μM/mM creatinine) (Bouatra et al., 2013).

1.5.4.2 Osmolality normalisation

This method is widely adopted in urine metabolomics studies as it is a direct measure of total dissolved solids in urine and is not subject to temperature and pressure changes (Warrack et al., 2009, Boudonck et al., 2009). Measurements are taken by observing the freezing point depression of samples relative to pure water. Osmolality normalisation has been shown to significantly reduce biological replicate variation in a study investigating metabolic responses to high and low doses of a phospholipidotic drug in male and female rats (Warrack et al., 2009).

1.5.4.3 Mass spectral signal normalisation

Unlike creatinine and osmolality this technique requires no additional measurements to be carried out and is achieve purely post data acquisition. Two methods exist for this normalisation, the first is mass spectrometer total signal (MSTS) and mass spectrometer total useful signal (MSTUS). The first MSTS normalises all mass spectral peak intensities to the sum peak intensity for the entire chromatogram, effectively giving all chromatograms a uniform sum peak intensity (Ganti and Weiss, 2011). This is becoming increasingly popular along with osmolality as a form of normalisation of urine sample data sets (Ganti and Weiss, 2011).

MSTUS on the other hand normalises to the sum peak intensity of peaks that are present in all samples in the batch. Following this normalisation technique, the sum intensity of these common peaks is the same for each chromatogram as opposed to the total peak intensities being equal (Warrack et al., 2009, Mattarucchi and Guillou, 2012). This eliminates the introduction of any bias due to large peaks associated with uncommon peaks such as those attribute to pharmaceuticals, diet and other xenobiotics (Mattarucchi and Guillou, 2012).

1.5.5 Importance of use of urine metabolomics studies in disease to date

The use of urine characteristics to diagnose or monitor a patient's health is not new, as both Hippocrates and Galen noted the potential of urine as a diagnostic biofluid. Indeed for over 6000 years, the colour, smell and taste of urine has been used to diagnose disease such as diabetes (Ryan et al., 2011). The use of urine metabolomics in the past 10 years has been used to detect metabolite profiles associated with disease (Gika et al., 2014, Mamas et al., 2011) and exposure to dietary compounds (Llorach et al., 2012, Wishart, 2008, Xie et al., 2013), pharmaceuticals (Lindon et al., 2007, Shockcor and Holmes, 2002) and environmental exposure (Johnson et al., 2012b, Rappaport, 2011). In addition to human health, urine from experimental animals are often analysed to determine health effects that can be used to infer human health outcomes. These experimental animals include disease models, or form part of exposure studies to pharmaceuticals or environmental contaminants where human studies are not practical.

1.5.5.1 Clinical studies utilising urine metabolomics

The use of metabolomic profiling of control and diseased patients offers and exciting prospect for further understanding metabolic responses to disease or biomarker discovery for; early diagnosis, assessing treatments, and finding potential targets for new drug therapies. Urine metabolomics has been for the most part implemented in research for cancers of the; kidney (Ganti and Weiss, 2011, Ganti et al., 2012a, Ganti et al., 2012b), bladder (Issaq et al., 2008b, Pasikanti et al., 2013, Huang et al., 2013), breast (Serkova et al., 2007, Slupsky et al., 2010), bowel (Dawiskiba et al., 2014), lung (Yang et al., 2010, Carrola et al., 2011), liver (Wu et al., 2009), prostate (Sreekumar et al., 2009), adrenal gland (Arlt et al., 2011) and several others. In several of these studies metabolite identification was not carried out as these were small pilot studies to demonstrate metabolomics can differentiate between cancer and control patients (Kind et al., 2007, Kim et al., 2011, Issaq et al., 2008b). Where metabolites were identified they highlighted the same metabolic pathways in all cancer types. The affected pathways were amino acid metabolism in particular tryptophan, Krebs cycle metabolites and lipid metabolism (Ganti and Weiss, 2011, Issaq et al., 2008b, Slupsky et al., 2010, Peng et al.,

2014). This is suggestive of changes in energy metabolism within the tumour which utilises glycolysis as an energy source thus depleting sources of Krebs cycle metabolites elsewhere in the body (Slupsky et al., 2010). However, none of these studies have lead to a unique biomarker or set of biomarkers unique to one particular cancer. The same lack of specificity while achieving metabolomic differentiation from controls is also observed in a number metabolomic studies including neurodegenerative diseases (Gebregiworgis et al., 2013, Michell et al., 2008), depression (Zheng et al., 2013), diabetes (Zhao et al., 2010, Connor et al., 2010), cardiovascular disease (Zhang et al., 2009), hepatitis (Wang et al., 2012b), inflammatory diseases (Dawiskiba et al., 2014, Lin et al., 2009), and obesity (Wang et al., 2011).

Metabolomics research in pharmaceutical development play roles in understanding the mechanism of action, progression of treatment and effect of toxicity (Shockcor and Holmes, 2002). Urine metabolomics is of particular interest in drug toxicity and overdose analysis. It has been demonstrated that different drugs have different metabolic effects in their target toxicity organs, for example acetaminophen, methotrexate and carbon tetrachloride all affect the liver in different ways. A metabolomic approach to hepatotoxicity in mice found each of these drugs induces its own metabolic changes giving each one a unique set of biomarkers. In each case, changes in urinary metabolite profiles preceded the increase in conventional serum aspartate aminotransferase, alanine aminotransferase and blood urea nitrogen levels typically used as markers of hepatotoxicity (Kumar et al., 2012). Acetaminophen (paracetamol) accounted for 40% of all liver injuries in the United States of America in 2005 (Loo et al., 2012). Due to this fact, acetaminophen toxicity has been widely investigated. In these studies, metabolomic analyses of urine revealed that antioxidants such as trigonelline, ferulic acid and S-adenosyl-L-methionine levels all decreased after acetaminophen exposure indicating oxidative damage (Sun et al., 2008). In addition, metabolites from the bile acid, energy and amino acid metabolic pathways were found to be deregulated following acetaminophen induced hepatotoxicity (Sun et al., 2012). The kidney is another major organ to be damaged due to pharmaceutical intake and several studies have investigated pharmaceutical induced nephrotoxicity using a urine metabolomics approach. The chemotherapeutic agent cisplatin (Portilla et al., 2006, Wen et al., 2011) and antibiotic gentamicin (Sun et al., 2012) have both been investigated in rats for nephrotoxicity and a range of biochemical pathways were shown to be disrupted including amino acid and catecholamine metabolism. These metabolic changes were again detectable before conventional serum tests showed signs of nephrotoxicity.

1.6 Human immunodeficiency virus (HIV)

The HIV lentivirus was first isolated in 1983, and is the cause of acquired immunodeficiency (AIDS) which had first been described in 1981. Since then over 60 million people have become infected and 25 million have died from the infection (Sharp and Hahn, 2011). HIV typically presented in previously healthy young men who have sex with men, or intravenous drug users. Later investigations identified heterosexual activity, blood transfusions and childbirth of HIV infected mothers as routes of infection (Levy, 1993). Prior to its isolation and identification, infection with HIV typically presented with emaciation, an onslaught of opportunistic infections and rare carcinomas such as Kaposi's sarcoma (Sharp and Hahn, 2011). When these symptoms are present in a patient they are said to have AIDS as a result of high levels of virus present in their blood and very low levels of CD4 positive (CD4+) T-lymphocytes at concentrations of <200 cells per μL of blood. This rapid loss of CD4+ cells is caused by a combination of factors, of which the main three are; the direct killing of infected cells, increased rate of apoptosis of infected cells and the increased immunological response of the blood CD8 cytotoxic cells that recognise infected cells (Levy, 1993, Pantaleo et al., 1993a).

1.6.1 HIV pathogenesis

Following the primary infection with HIV, the virus rapidly replicates and disseminates to the lymphoid organs during seroconversion/acute HIV infection. During the body's cellular and humoral response viral replication is inhibited within weeks of primary infection, leading to a chronic and persistent infection for a number of years during clinical latency. This subsequently leads to an advanced clinical disease with a very high level of mortality frequented by opportunistic infections and cancers (Figure 1.6) (Fauci, 1996, Pantaleo et al., 1993a).

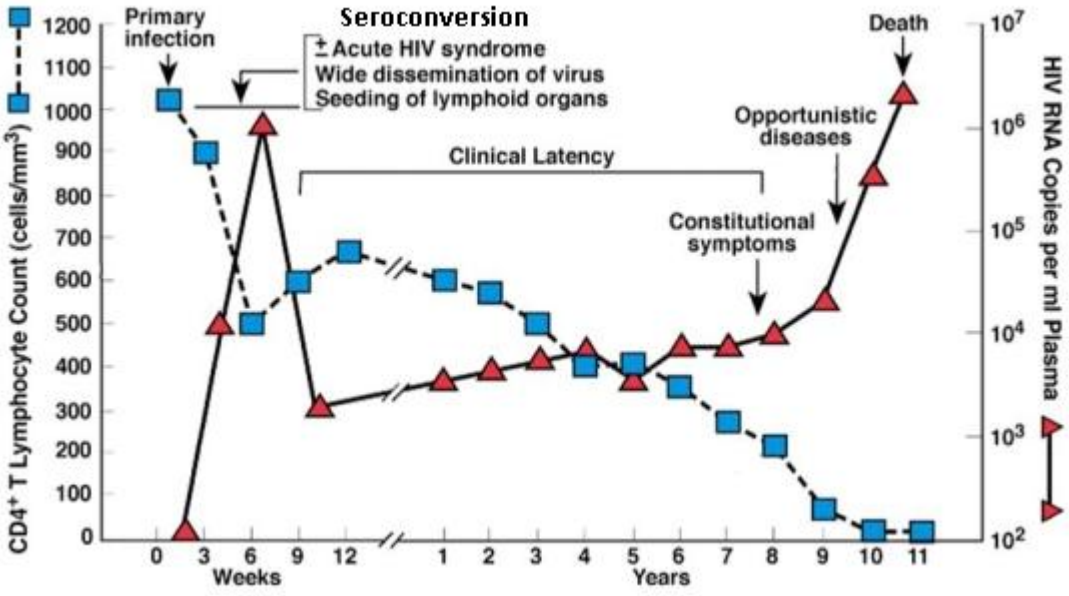


Figure 1.6: Typical progression of HIV infection

Primary infection is followed by rapid replication and dissemination during which viral load peaks while the CD4 count reaches a nadir. Following the body's response clinical latency follows for several years during which HIV remains as a chronic infection. Following this, advanced clinical HIV develops during which AIDS presents as severe opportunistic infections cancer and eventual death. (Adapted from: Host factors and the pathogenesis of HIV-induced disease) (Pantaleo et al., 1993a)

HIV is a positive sense strand RNA retrovirus and the main cellular targets of HIV are CD4 membrane antigen positive cells. These are typically the white blood cells; T lymphocytes and monocytes, although glial cells, macrophages and dendritic cells are also infected (Haseltine, 1991, Greene, 1991). The viral envelope of the HIV virus contains a number of proteins including glycoprotein (gp) 120 and gp41 which are viral proteins and also major histocompatibility complexes class I and II which are acquired from host cells during viral budding. Within the viral capsid reside the two strands of RNA which contains the 9 HIV genes and also within the capsid are the enzymes reverse transcriptase, protease and integrase which are utilised in the replicative process (Levy, 1993).

HIV enters its target cells via interaction of gp120 with the CD4 receptor for which is has a very high affinity, this is facilitated by gp41. Following the fusion of the viral envelope with the cell membrane the HIV capsid is released into the cell (Gomez and Hope, 2005). The viral RNA and enzymes translocate to the nucleus where the reverse transcriptase transcribes the RNA into double stranded DNA. Integrase subsequently integrates the HIV DNA into a host chromosome as a provirus (Levy, 1993). Once integrated into the host chromosome the virus can remain latent for several years leading to clinical latency. Once the provirus becomes activated the host RNA polymerase translates the HIV genome and the messenger RNA is transported to the cytoplasm where it is transcribed into HIV proteins and enzymes. Once

assembled, the viroid translocates to the cell membrane and buds through the cell membrane during which the viroid acquires a membrane coating from the host cell on to which viroid proteins are then expressed. The budding and release of virus into the extracellular matrix is mediated by the viroid protease (Greene, 1991).

1.6.2 HIV treatments

HIV is a chronic condition with no cure yet available, however the viral load and AIDs related illness can be suppressed with antiretroviral therapy (ART). There are a number of classes of antiretrovirals (ARVs) in use each of which acts via a different mechanism and each attacks the virus at different points of its replication cycle. Nucleoside reverse transcriptase inhibitors (NRTIs) and non nucleoside reverse transcriptase inhibitors (NNRTIs) both inhibit viral reverse transcriptase preventing the formation of viral DNA. The former acts by competitively binding with the active site of the reverse transcriptase, while the latter is non competitive and inhibits the enzyme via allosteric binding elsewhere on the protein (Menéndez-Arias, 2002). Protease inhibitors (PI) limit viral replication via the competitive inhibition of HIV protease by binding to the active site (Menéndez-Arias, 2002). Integrase inhibitors (IHI) are used to inhibit the integration of HIV DNA into host cell chromosomes. These four classes make up the majority of ARVs in use to date. Current guidelines stipulate patients should receive a combination of ART (cART) in order to reduce viral load and transmission. The current recommendation is that two NRTIs and a ritonavir-boosted protease inhibitor or an NNRTI and integrase inhibitor are used (Williams et al., 2014). Since the introduction of antiretroviral therapy patients have experienced a longer life span, and overall AIDS defining condition mortality has dropped, whilst age related mortality has risen from causes such as cardiovascular disease (Sabin, 2013, Marin et al., 2009, Deeks and Phillips, 2009). Due to the extended life time of these patients, they are now exposed to ARVs over a prolonged period of time. Many patients report metabolic side affects which are associated with ART and these symptoms include insulin resistance, lipidystrophy, hyperlipidemia and abdominal obesity (Jain et al., 2001). In addition, more serious side effects are associated with particular drugs. Tenofovir, an NRTI, is associated with nephrotoxicity and the primary target for toxicity is the renal proximal tubules, and in some patients this can lead to Fanconi syndrome (Hall et al., 2011a). The mode of action for tenofovir toxicity is thought to be abnormalities in the mitochondria of the renal proximal tubules (Kohler et al., 2009, Hall et al., 2011a). All ART is associated with hepatotoxicity and in particular in patients co-infected with hepatitis, leading to liver cirrhosis and transaminase elevation (Rivero et al., 2007). Atazanavir, a protease inhibitor, in particular is associated with increased levels of bilirubin due to the

inhibition of uridine 5'-disphospho-glucuronosyltransferase (Pineda et al., 2006). As many metabolomic studies have been successful in uncovering the pathways associated with toxicity of other pharmaceuticals, it is possible that it may also elucidate mechanisms of cART toxicity and underlying metabolic changes associated with long term cART exposure (Sitole et al., 2013).

1.6.3 HIV infection and metabolism

The metabolic effect of HIV infection had been observed by indirect calorimetry in HIV positive patients who were shown to expend more energy in a resting state due to the greater energy demands of infected cells (Hommes et al., 1990, Hommes et al., 1991, Kosmiski, 2011). These patients have also been shown to oxidise fatty acids at an increased rate relative to controls (Hommes et al., 1991). The effect of resting energy expenditure however returns to normal when the viral load is under control due to anti-retroviral treatment (Kosmiski, 2011). In addition, the metabolomic activity in the brain has also been observed to be greater in asymptomatic HIV positive patients (Pascal et al., 1991). Metabolomic studies have revealed that HIV infection can cause malnutrition due to mitochondrial dysfunction, altered lipid protein and carbohydrate metabolism and changes in fat distribution (Sitole et al., 2013).

Despite being one of the most well researched viruses metabolomic analysis of HIV infection has remained very limited. To date metabolomic studies of HIV have focussed mainly on blood products such as plasma (Cassol et al., 2013) and serum (Williams et al., 2012, Philippeos et al., 2009) or saliva (Ghannoum et al., 2013). Several studies have looked at the neurological implications of HIV infection by investigating the cerebrospinal fluid (CSF) metabolome (Wikoff et al., 2008, Cassol et al., 2014), and a further study investigated metabolomic altercations in CD4 expressing cells of the immune system (Hollenbaugh et al., 2011). In all these studies metabolic deregulation of amino acid, carbohydrate and lipid metabolite pathways was observed as a result of the increased metabolic rate of the HIV infected cells. A flaw of several of these studies is that it is not clear if the metabolites detected were a result of disease progression as both CD4 count and viral loads were not reported, or when they were reported the values exhibited a wide range (Sitole et al., 2013).

These studies suggest that metabolomic studies of HIV in biofluids are capable of detecting metabolite variations as a result of HIV infection and cART intervention. Metabolomic analysis of urine may discover new markers of HIV infection and/or confirm that those markers detected in other biofluids are also present in urine. This would be advantageous as not only is urine a non-invasive biofluid, it is also sterile meaning there is no

risk of infection for the researcher. Using more sensitive analytical techniques further markers associated with HIV infection may be uncovered which may aid in determining progression of the disease and the efficacy of pharmaceutical intervention.

1.7 Aims and objectives

Urine is a very complex biological fluid, which offers insight to the body's endogenous metabolic pathways and exposure to xenobiotics making it an ideal matrix for metabolomic analysis for disease states. However improvements in sensitivity are needed in order to cover less abundant components of the metabolome such as signalling compounds important in disease processes.

The aims of this PhD thesis are 4 fold:

1. The development of new MS methodologies using nanoUHPLC-nanoESI-TOFMS to enhance the overall coverage of the urinary metabolome
2. To improve urine sample preparation methodology prior to MS analysis
3. The development of new data handling protocols to eliminate unnecessary metabolome variation caused by diet, pharmaceutical intake and normalisation biases
4. To implement these techniques in a HIV cohort metabolomics study to investigate HIV pathogenesis and metabolic consequences of cART intervention.

The work to meet these aims is detailed in the following 5 chapters:

- Chapter 2 describes preliminary method development and assessment of nUHPLC-nESI-TOFMS platform for metabolomic analysis and includes a consideration of:
    - Nanoflow column selection
    - Gains in sensitivity relative to conventional techniques using a standard test chemical mixture
    - The reliability of the nUHPLC-nESI-TOFMS platform using urine as a sample matrix for metabolomics analysis

The work presented in this chapter has been written up and published as a application note in the Journal of Mass Spectrometry (Chetwynd et al., 2014).

- Chapter 3 demonstrates the improved coverage of the urinary metabolome utilizing new sample preparation methodologies and compares nUHPLC-nESI-TOFMS with a conventional UHPLC-ESI-TOFMS platform for urinary metabolomics. The work comprises:

- The development of SPE methodology and a comparison with the use of traditional sample preparation techniques for urine.
- Metabolomic analysis of urine samples extracted by SPE on nUHPLC-nESI-TOFMS and conventional UHPLC-ESI-TOFMS to assess gains in metabolomic coverage using the nano platform.

This chapter has been published as a research article in Analytical Chemistry (Chetwynd et al., 2015)

- Chapter 4 details a study into a comparison of normalisation techniques for processing data sets from test urine samples extracted by SPE and analysed by nUHPLC-nESI-TOFMS and comprises:
  - An assessment of different normalisation techniques employed prior to or post nUHPLC-nESI-TOFMS analyses.
  - The use of a mass exclusion list to remove confounding xenobiotics from data sets
- Chapters 5 and 6 comprise case studies from a HIV cohort utilizing methods developed in chapters 2-4.
- Chapter 7 comprises of a general discussion of the thesis work and examines future developments needed in development of data acquisition and processing for urinary metabolomics.

Chapter 2: Evaluation of analytical performance and reliability of direct nanoLC-nanoESI-high resolution mass spectrometry for profiling the (xeno)metabolome

2.1 ABSTRACT

Mass spectrometry profiling techniques are used for analysing metabolites and xenobiotics in biofluids, however detection of low abundance compounds using conventional MS techniques is poor. To counter this, the use of nanoflow ultra-high-pressure liquid chromatography-nanoelectrospray ionization-time-of-flight MS (nUHPLC-nESI-TOFMS), which has been used primarily for proteomics, was assessed as an innovative prospect for sensitive analysis of small molecules. This study revealed that compared to conventional UHPLC-ESI-TOFMS, nUHPLC-nESI-TOFMS enhanced instrumental detection limits of a variety of metabolites and xenobiotics by between 2 and 2000 fold. In addition, this study demonstrates for the first time excellent repeatability and reproducibility for analysis of urine samples using nUHPLC-nESI-TOFMS, supporting implementation of this platform as a novel approach for high-throughput (xeno)metabolomics.

2.2 Introduction

Metabolomic analyses consist of profiling of the many endogenous metabolites present in biological matrices (i.e., the metabolome), in order for example, to discover potential biomarkers of disease or toxicant exposure (Kell, 2006). In addition, xenobiotics and their metabolic by-products can also be analysed in the same samples, and these compounds are referred to as the xenometabolome (Al-Salhi et al., 2012, Holmes et al., 2007, Nicholson, 2006). Analysis of the (xeno)metabolome may increase scientific understanding of the impact of toxicants, pharmaceuticals and lifestyle factors on human health (Bonvallot et al., 2014, Bouhifd et al., 2013). The use of UHPLC-ESI-TOFMS enables the detection of polar to apolar (xeno)metabolites at high mass resolution, and has been extensively used for profiling of small organic molecules in sample extracts (Gika et al., 2014, Theodoridis et al., 2012). However, using this technique, the detection of very low abundance metabolites such as signalling compounds or chemical contaminants is limited as they can suffer from poor ionization efficiency or ion suppression from co-eluting metabolites (Gika et al., 2014). As a result, more efficient separation and ionization techniques are required to undertake profiling of trace level metabolites in sample extracts.

In recent years, new technological advances have paved the way for nano scale chemical analysis in the form of nanoflow UHPLC-nanoESI-MS (nUHPLC-nESI-MS) (Wilm and Mann, 1996, Berggren et al., 2002, Gangl et al., 2001, Juraschek et al., 1999, Schmidt et al., 2003). The key advancement of this technique is the improved sensitivity as a result of using

lower flow rates and nanospray emitters (Wilm and Mann, 1996). The mechanisms for improved sensitivity as a result of nanospray ionization have been detailed in Chapter 1 Section 1.3.2.2.1.

To date, nUHPLC-nESI-MS techniques have been used primarily for proteomic analyses (Liu et al., 2008, Emmett and Caprioli, 1994, Shevchenko et al., 1996) and have been for the most part overlooked for metabolomics applications (Medina et al., 2013b). The increased ionization efficiency of nanoscale systems offers significant improvements for small molecule analysis and is now starting to be implemented for some non-targeted applications. However, many of the nanoflow systems described use traditional LC setups and split the flow prior to entry to the MS. Compared with split flow, direct injection into nanoflow LC could enhance sensitivity by reducing sample loss, and should result in more reproducible nanoflow rates into the source (Gama et al., 2013). Therefore, the use of direct nanoflow coupled with nESI-MS could be advantageous to improve the detection of low levels of many biologically important signalling compounds during metabolite profiling of extracts of tissue or biofluids. These compounds include, for instance, estrogenic steroids which are poorly ionized by ESI, or prostanoids present at pg/mL concentrations in blood plasma and urine (Wishart et al., 2013), or some pharmaceuticals and other xenobiotics which can often accumulate as low concentrations of complex mixtures in biological samples with potential consequences for human health (Melzer et al., 2010, Rochester, 2013, Daughton and Ruhoy, 2008).

To date, no studies are present that investigate the benefits associated with nUHPLC-nESI-TOFMS for metabolomics. Of particular importance is the repeatability and reproducibility of the peak area and retention time, these properties are essential to the peak picking process in the analysis of metabolomic data.

2.2.1 Study aims

The aim of this study was to evaluate the detection limits, repeatability and reproducibility of a direct nUHPLC-nESI-TOFMS method to perform high throughput (xeno)metabolomic analyses. To achieve this, the effects of nano column chemistries and porosities on chromatographic separation were investigated using a mix of 78 compounds including endogenous metabolites and xenobiotics. The detection limits of the nanoflow-nanospray platform were then compared to conventional UHPLC-ESI-TOFMS. Finally, spiked replicate urine samples were analysed to assess the repeatability and reproducibility of the nUHPLC-nESI-TOFMS method for (xeno)metabolomic analyses.

2.3 Experimental

2.3.1 Chemicals

UHPLC grade solvents were purchased from Fisher Scientific (Walkerburn, Scotland, UK). Strata X-AW 60 mg/3ml solid phase extraction (SPE) cartridges were purchased from Phenomenex (Macclesfield, U.K). Deuterated compounds were used as internal standards (IS); 17β-estradiol 2,4,16,16-d4 sodium 3-sulfate (E2-d4-S, >99% D atom), carbamazepine (ring-d10), venlafaxine (N,N-dimethyl-d6), and diclofenac (phenyl-d4) were purchased from Cambridge Isotope Laboratories Inc. (MA, USA). Progesterone-2,2,4,6,6,17R,21,21,21-d9 (P-d9, 98% D atom) was purchased from CDN isotopes (Quebec, Canada) and prostaglandin E2-d4 (9-oxo-11α,15S-dihydroxy-prosta-5Z,13E-dien-1-oic-3,3,4,4-d4 acid) was purchased from Cayman Chemical Company (MI, USA). All other standards and reagent chemicals were purchased from Sigma-Aldrich Company Ltd., Dorset, U.K.

2.3.2 Standard preparation

The test mixture was chosen to include neutral, acid and basic compounds with an octanol-water partition coefficient (log Kow) range of -2.51 to 6.3. Log Kow values were obtained from ChemSpider (http://www.chemspider.com) or using ACD/LABS V12.0 (Advanced Chemistry Development, Berks, UK). Details of the 78 metabolites and xenobiotics (49 detected in positive ESI and 29 in negative ESI) are given in Table 2.1. A stock solution (1 mg/ml) of the mixture was prepared in methanol (MeOH), and subsequently diluted to obtain a final concentration of 0.2 µg/ml in 50/50 MeOH/water (v/v) which was used for column comparison and analytical limit of detection analysis. These compounds were chosen as they represent a range of metabolic pathways or xenobiotics typically present at trace levels in the urine. The more common high abundance metabolites were not chosen as current urine metabolomic methods offer highly reliable and reproducible methods for their analysis. The focus of this study is to develop a methodology ideally suited to low abundance metabolites typically not detected in current untargeted metabolomic urine methodologies.

Table 2.1: Identity of the 78 standards in the standard mixture, and their molecular formula, log Kow, theoretical ion mass and retention time (Rt)

Compound	Formula	Kow	Theoretical mass	Rt	Compound	Formula	Kow	Theoretical mass	Rt
Neurotransmitters					**NSAID's**				
Epinephrine	$C_9H_{13}NO_3$	-0.69	M+H 184.0974	3.01	Diclofenac	$C_{14}H_{11}Cl_2NO_2$	4.02	M-H 294.0089	21.64
Dopamine	$C_8H_{11}NO_2$	0.38	M+H 154.0868	3.34	Ibuprofen	$C_{13}H_{18}O_2$	3.97	M-H 205.1229	22.01
Serotonin	$C_{10}H_{12}N_2O$	0.79	M+H 177.1028	3.38	Naproxen	$C_{14}H_{14}O_3$	3.1	M-H 229.0865	16.30
Melatonin	$C_{13}H_{16}N_2O_2$	1.65	M+H 233.1290	9.81	Ketoprofen	$C_{16}H_{14}O_3$	3	M+H 256.0865	15.41
5-Hydroxyindoleacetic acid	$C_{10}H_9NO_3$	1.01	M+H 192.0661	7.31	Mefenamic acid	$C_{15}H_{15}NO_2$	5.12	M+H 242.1181	24.43
Nucleotides					Sulfasalazine	$C_{18}H_{14}N_4O_5S$	3.81	M-H 397.0607	12.08
Cytidine	$C_9H_{13}N_3O_5$	-2.51	M+H 244.0933	3.38	Indomethacin	$C_{19}H_{16}ClNO_4$	4.27	M-H 356.0690	21.85
Cytosine	$C_4H_5N_3O$	-1.48	M+H 112.0511	3.48	Paracetomol	$C_8H_9NO_2$	0.27	M+H 152.0712	3.41
Amino acids					**Lipid lowering agents**				
Histidine	$C_6H_9N_3O_2$	-3.32	M+H 156.0773	3.94	Clofibric acid	$C_{10}H_{11}ClO_3$	2.84	M-H 213.0318	15.37
Tryptophan	$C_{11}H_{12}N_2O_2$	1.22	M+H 205.0977	3.84	Gemfibrozil	$C_{15}H_{22}O_3$	4.3	M-H 249.1491	25.07
Creatinine	$C_4H_7N_3O$	-1.77	M+H 114.0667	3.36	Bezafibrate	$C_{19}H_{20}ClNO_4$	4.25	M+H 362.1159	16.90
Vitamins					Pravastatin	$C_{23}H_{36}O_7$	3.1	M-H 423.2383	11.77
Nicotinic acid	$C_6H_5NO_2$	0.69	M+H 124.0399	3.95	Simvastatin	$C_{25}H_{38}O_5$	4.68	M+H 419.2797	28.06
Retinoic acid	$C_{20}H_{28}O_2$	6.3	M+H 301.2168	20.52	**Antiplatelet**				
Folic acid	$C_{19}H_{19}N_7O_6$	-2.81	M+H 442.1475	6.95	Dipyridamole	$C_{24}H_{40}N_8O_4$	-1.22	M+H 505.3251	10.77
Steroids					Clopidogrel	$C_{16}H_{16}ClNO_2S$	3.82	M+H 322.0669	22.94
Cortisol	$C_{21}H_{30}O_5$	1.62	M+H 363.2171	10.91	Warfarin	$C_{19}H_{16}O_4$	2.23	M-H 307.0970	18.47
Cortisone	$C_{21}H_{28}O_5$	1.81	M+H 361.2015	11.12	**SSRI and antipsychotics**				
Corticosterone	$C_{21}H_{30}O_4$	1.99	M+H 347.2222	12.90	Carbamazepine	$C_{15}H_{12}N_2O$	2.25	M+H 237.1028	11.92
11-ketotestosterone	$C_{19}H_{26}O_3$	1.92	M+H 303.1960	11.85	Fluoxetine	$C_{17}H_{18}F_3NO$	3.93	M+H 310.1419	14.00
21-Hydroxypregnenolone	$C_{21}H_{32}O_3$	3.75	M+H 333.2430	16.82	Paroxetine	$C_{19}H_{20}FNO_3$	3.7	M+H 330.1505	12.26
17α-Hydroxyprogesterone	$C_{21}H_{30}O_3$	3.08	M+H 331.2273	16.83	Sertraline	$C_{17}H_{17}Cl_2N$	5.29	M+H 306.0816	14.37
Androstenedione	$C_{19}H_{26}O_2$	2.76	M+H 287.2011	17.62	Venlafaxine	$C_{17}H_{27}NO_2$	3.28	M+H 278.2120	9.09
Progesterone	$C_{21}H_{30}O_2$	3.83	M+H 315.2324	23.99	Clozapine	$C_{18}H_{19}ClN_4$	3.35	M+H 327.1376	9.64
Pregnenolone	$C_{21}H_{32}O_2$	3.89	M+H 317.2481	23.65	**β-Blockers**				
Testosterone	$C_{19}H_{28}O_2$	3.27	M+H 289.2168	15.77	Metoprolol	$C_{15}H_{25}NO_3$	1.63	M+H 268.1913	7.96
Testosterone glucuronide	$C_{25}H_{36}O_8$	2.51	M-H 463.2332	11.50	Propranolol	$C_{16}H_{21}NO_2$	2.9	M+H 260.1651	10.00
Estrone	$C_{18}H_{22}O_2$	3.43	M-H 269.1542	17.65	Atenolol	$C_{14}H_{22}N_2O_3$	0.16	M+H 267.1701	3.31
Estrone glucuronide	$C_{24}H_{30}O_8$	1.58	M-H 445.1863	10.97	**Pesticides-fungicides-antibacterial**				
Estrone sulfate	$C_{18}H_{22}O_5S$	0.95	M-H 349.1110	12.30	Diazinon	$C_{12}H_{21}N_2O_3PS$	3.81	M+H 305.1089	25.67
Estradiol	$C_{18}H_{24}O_2$	3.94	M-H 271.1699	15.45	Piperonyl butoxide	$C_{19}H_{30}O_5$	4.29	M+H 339.2171	23.63
Estradiol glucuronide	$C_{24}H_{32}O_8$	3.18	M-H 447.2019	10.08	Terbutryn	$C_{10}H_{19}N_5S$	3.7	M+H 242.1439	15.03
Eicosanoids					Atrazine	$C_8H_{14}ClN_5$	2.82	M+H 216.1016	13.17
Prostaglandin E2	$C_{20}H_{32}O_5$	3.52	M-H 351.2071	14.02	Miconazole	$C_{18}H_{14}N_2OCl_4$	6.25	M+H 414.9938	20.41

Table 2.1 continued									
Prostaglandin B2	$C_{20}H_{30}O_4$	3.4	M-H 333.2066	18.30	Propiconazole	$C_{15}H_{17}Cl_2N_3O_2$	3.72	M+H 342.0776	22.72
Thromboxane B2	$C_{20}H_{34}O_6$	2.77	M-H 369.2277	12.50	Triclosan	$C_{12}H_7Cl_3O_2$	4.78	M-H 286.9433	27.00
Fatty acid metabolites					Chlorophene	$C_{13}H_{11}ClO$	4.18	M-H 217.042	23.65
Arachidonic acid	$C_{20}H_{32}O_2$	8.07	M-H 303.2324	32.15	**Endocrine disruptors/metabolites**				
Sphingosine	$C_{18}H_{37}NO_2$	5.53	M+H 300.2903	23.40	Bisphenol A	$C_{15}H_{16}O_2$	3.64	M-H 227.1072	15.12
Bile acids					Butylparaben	$C_{11}H_{14}O_3$	3.47	M-H 193.0865	17.38
Taurocholic acid	$C_{26}H_{45}NO_7S$	0.01	M-H 514.2838	11.81	Ethinylestradiol	$C_{20}H_{24}O_2$	3.67	M-H 295.1698	17.60
Cholic acid	$C_{24}H_{40}O_5$	3.52	M-H 407.2797	17.09	Ethinylestradiol glucuronide	$C_{26}H_{32}O_8$	2.21	M-H 471.2019	11.10
Deoxycholic acid	$C_{24}H_{40}O_4$	3.5	M-H 391.2848	23.30	Ethinylestradiol sulfate	$C_{20}H_{24}O_5S$	1.77	M-H 375.1266	12.07
Food additives					**Flame retardants**				
Caffeine	$C_8H_{10}N_4O_2$	0.16	M+H 195.0882	7.12	Triphenyl phosphate	$C_{18}H_{15}O_4P$	4.59	M+H 327.0786	26.07
					Tris(2-butoxyethyl)phosphate	$C_{18}H_{39}O_7P$	3.75	M+H 399.2512	27.12

Rt were obtained using a HSS-T3 100 Å column on nUHPLC-nESI-TOFMS with 0.01% formic acid modified water and acetonitrile as mobile phase A and B, respectively, at a flow rate of 700 nL/min. Log Kow values were obtained from ChemSpider (http://www.chemspider.com) or ACD/LABS V12.0. NSAID refers to non-steroidal anti-inflammatory drugs and SSRI to selective serotonin reuptake inhibitors.

2.3.3 Construction of in-house nanospray emitters

Nanospray emitters were prepared in-house using polyetheretherketone (PEEK) tubing (100 µm I.D.), cut to a length of 6 cm (Kinesis, St Neots, UK). A fused silica tube, 10 µm I.D. and 90 µm external diameter (Waters, Manchester, UK), was cut to 8 cm length and passed through the PEEK tubing. Following construction, emitters were sonicated in HPLC grade MeOH for 10 min prior to use.

2.3.4 nUHPLC-nESI-TOFMS analysis and comparison of column performance

The nUHPLC-nESI-TOFMS experiments were performed on a Waters nanoAcquity UHPLC linked to a Waters Xevo G2 TOFMS, equipped with a nESI source (Waters, Manchester, U.K). The Xevo G2 TOFMS was tuned to a mass resolution of 20,000 with a mass spectra range of 50-1000 *m/z*. A capillary voltage of 3.5 kV and cone voltage of 30 V was used in both nESI modes, with collision energy of 6 eV and a nitrogen flow of 60 l/hr for the cone and 300 l/hr for desolvation. A leucine enkephalin lockspray standard in 1/1 MeOH/water (v/v) (2 ng/µl) was infused at 700 nl/min in both ionization modes. Standard solutions or samples (0.5 µl) were loaded directly onto column with a partial loop injection. The sample loop was rinsed successively with 1500 µl of acetonitrile (ACN) and water after each injection to avoid sample cross-contamination. Three nUHPLC columns were compared: Waters nanoAcquity UHPLC Ethylene Bridged Hybrid (BEH) C18 100 µm x 100 mm x 1.7 µm, 300 Å; BEH C18 100 µm x 100 mm x 1.7 µm, 130 Å; and High Strength Silica (HSS) T3 100 µm x 100 mm x 1.8 µm, 100 Å (Waters, Manchester, UK). The chemistries of these columns differ in their back bone; the BEH is an ethylene-silica hybrid whereas the HSS is 100% silica. The differences between the C18 and T3 chemistries is their density within the column, the T3 trifunctional C18 alkyl chains are less dense than the C18 making it better suited for polar compounds (Neue et al., 2006). The columns were maintained at 25 °C. UHPLC grade water was used as solvent A and ACN as solvent B, both modified with 0.01% formic acid (FA). The flow rate was 700 nl/min with a gradient of 0-4 min from 10 to 30% B, 4-18 min from 30% to 50% B, 18-30 min from 50 to 100% B, 30-35 min 100% B, and equilibration to initial conditions in 15 min. For all tested conditions, injections were performed in triplicate with each standard at 100 pg on column. Retention times and integrated peak areas were generated using QuanLynx MassLynx V4.1 software. A two-way analysis of variance (ANOVA) with Tukey post hoc analysis was used to calculate significant variation in between mean peak areas (n=3) for each of the 78 compounds across the three columns tested. ANOVAs were performed using GraphPad Prism V6.0.

The column resolution, (*R*), was calculated using equation 2.1 (*Eq* 2.1) to quantify separation between peaks.

$$Eq\ 2.1. \quad R = \frac{2\left[Rt_2 - Rt_{ref}\right]}{PW_{ref} + PW_2}$$

Rt_{ref} and Rt_2 correspond respectively to the retention time of the reference compound and compound 2, and PW_{ref} and PW_2 the respective widths at half peak height. Column resolution was calculated using propranolol in positive mode, and testosterone glucuronide in negative) mode as early eluting reference peaks for each resolution calculation. For negative mode, standards used for calculations of column resolution were thromboxane B2, taurocholic acid, ibuprofen and arachidonic acid. In positive mode, testosterone, androstenedione, progesterone and tris(2-butoxyethyl)phosphate were used. These compounds were chosen because they elute throughout the entire chromatographic space.

2.3.5 Comparison of nUHPLC-nESI-TOFMS and UHPLC-ESI-TOFMS detection limits

The detection limits of the nUHPLC-nESI-TOFMS and conventional UHPLC-ESI-TOFMS platforms were compared by assessment of the instrumental limits of detections (IDL) of the standard mixture which were calculated for both systems. The IDLs were determined as the amount injected on column with a signal to noise ratio (S/N) = 3. The HSS T3 column was used to perform the comparison study for the nUHPLC-nESI-TOFMS. The conventional UHPLC comprised of a Waters Acquity UHPLC utilizing a UHPLC HSS-T3 column (50 mm x 1 mm x 1.8 µm), linked to a Micromass Q-TOFMS Ultima (Waters, Manchester, UK). The mobile phase and gradients were the same for both platforms (see details above for nUHPLC-nESI-TOFMS) but with a flow rate of 0.2 ml/min for UHPLC-ESI-TOFMS. Injection volumes were 5 µl for UHPLC-ESI-TOFMS analyses. The quantity of each standard injected on each platform ranged from 0.05 to 2000 pg.

2.3.6 Extraction of urine samples

Urine was collected from a healthy 23 year old male, with written informed consent and ethical approval from the University of Sussex Life Sciences and Psychology Ethics Committee. A total of 10 midstream first morning voids were collected and stored in 10% MeOH at -80 °C. An aliquot from each sample was combined to produce a pooled urine sample of 1 ml which was spiked with 500 ng of each deuterated IS and extracted on a Strata X-AW SPE cartridge. The SPE cartridge was primed with 1 mL MeOH, washed with 1 mL MeOH before loading with the urine samples. The cartridge was subsequently washed with 1 mL water and eluted with 1 mL 5% ammonium hydroxide in MeOH followed by 1 mL ethyl acetate. Eluates

were combined, dried under vacuum and reconstituted in 500 µl 50/50 MeOH/water. This sample preparation is further detailed and analysed in chapter 3.

2.3.7 Repeatability and reproducibility experiments

The intra-day repeatability of the nUHPLC-nESI-TOFMS method for both nESI modes was assessed by injecting 10 analytical replicates of the urine samples. The inter-day reproducibility was tested by injecting 10 analytical replicates of the composite extracts of urine samples on three separate days in both nESI modes. Coefficients of variation (CV) for the peak areas and retention times were calculated for each IS to assess the repeatability and reproducibility of the nUHPLC-nESI-TOFMS method. The analytical variability of the urine (xeno)metabolomes were also assessed based upon the method proposed by Want et.al for conventional UHPLC-ESI-TOFMS (Want et al., 2010). The response of each signal was normalised to the total ion signal of the sample using Waters MarkerLynx V4.1. The CV of the mean normalised response was calculated for all common peaks and the number of signals with a CV of less than 30% expressed as a percentage of the total.

2.4 Results and discussion

2.4.1 Column comparison

The effect of the column phase on the chromatography and the analyte response (i.e., peak area) of the mixture of 78 compounds were evaluated using three reversed phase nanoflow Acquity UHPLC columns with 2 column chemistries and 3 different pore sizes; BEH C18 300 and 130 Å, and HSS T3 100 Å. All 3 columns were investigated on the same day, and the performance of the MS was assessed by ensuring the signal intensity of the lockmass standard, leucine enkephalin, infused into the nESI varied by no more than 10% between analytical runs. Mean peak areas were used to assess analyte response, and the retention time and column resolution were used as a measure of chromatographic separation. ANOVA of the peak areas for each of the 78 compounds revealed that the response of many of the analytes was similar between the three column types (Table 2.2). However, from analysis in both nESI modes, the response of 33-34 of the standard compounds was significantly higher ($p \leq 0.05$) on either of the low porosity HSS T3 100 Å or BEH C18 130 Å columns compared with the higher porosity BEH-C18 300 Å column. A reduction in pore size increases the surface area of the stationary phase and subsequently can increase the retention and the rate of mass transfer between mobile and column phases (Motokawa et al., 2002). This may explain the increased retention, separation of selected analytes on the lower porosity columns (Table 2.3). The base peak intensity (BPI) chromatogram of the standard mixture analysed in positive nESI mode on the three columns is given in Figure 2.1. Analysis of the mixture in negative nESI mode did not

result in a clear BPI, and selected ion chromatograms of key standards analysed on the three columns is given in Figure 2.2. For these compounds, the additional retention observed for the HSS T3 100 Å and BEH C18 130 Å columns enabled compounds to elute in a greater quantity of ACN, which owing to its increased volatility relative to water is preferential for increased ionization efficiency of the analyte (New and Chan, 2008, Spagou et al., 2011). Hence for some analytes, the difference in pore size between the columns may explain the increased peak area observed with the low porosity columns.

Nevertheless, despite the advantages afforded by reduced pore size, the BEH C18 300 Å conferred a significant increase in mean peak area for some analytes including arachidonic acid, sulfalazine, melatonin and deoxycholic acid compared with the HSS T3 100 Å phase. In addition chlorophene, estradiol sulfate, estrone sulfate, testosterone glucuronide, estrone glucuronide and epinephrine had mean peak areas significantly greater on the BEH C18 300 Å compared to the BEH C18 130 Å phase. Many of the analytes showed poor peak shape on the lower porosity columns and tailed less on the BEH C18 300 Å. For example the column resolution of arachidonic acid was highest on the high porosity BEH column (Table 2.3).

Retention of many of the analytes was favoured by the HSS phase compared with the BEH phase. The HSS T3 utilizes silanol particles combined with a reduced density of trifunctional C18 alkyl chains compared to the BEH C18, therefore HSS T3 preferentially retains more polar and mid-polar compounds than the C18 columns (Neue et al., 2006). Additionally, the better column resolution observed with HSS may influence analyte responses due to less co-elution and subsequently less competition for ionization (Table 2.3). Increased separation of compounds in chemical analysis is a fundamental goal in metabolomics as it reduces ion suppression in complex mixtures and matrixes, thus increasing the overall performance of the system. It is noteworthy that the very polar compounds (i.e., epinephrine, dopamine, serotonin, cytidine, cytosine, histidine, nicotinic acid, and acetaminophen) are generally poorly retained on reversed phase chemistries and were not retained on any of the columns used in this study and were detected during the first minutes of the chromatogram (i.e. between 2-4 minutes) (Figure 2.1).

Table 2.2: Comparison of the peak areas of 78 (xeno)metabolites between nUHPLC column types

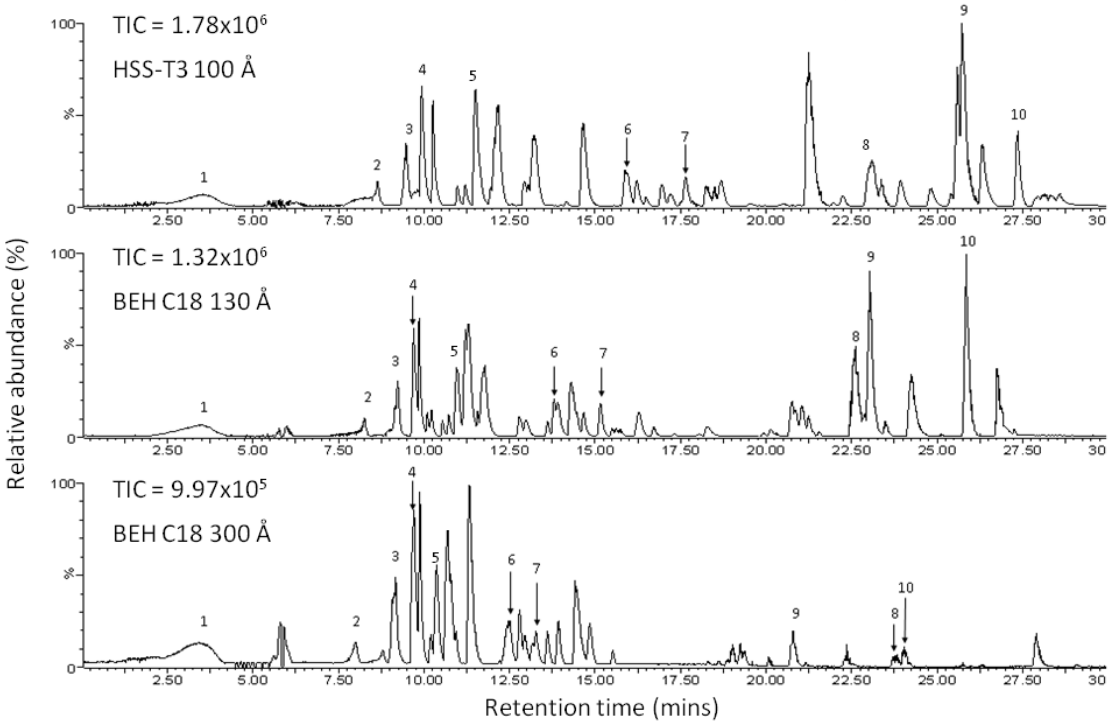
	HSS T3 100 Å	vs	BEH C18 300 Å	HSS T3 100 Å	vs	BEH C18 130 Å	BEH C18 130 Å	vs	BEH C18 300 Å
Number of compounds with a significantly higher peak area	34		4	19		11	33		10
Fold increase in peak area	*1.5 – 4.0*		*1.3 - 1.5*	*1.5 – 6.0*		*1.1 - 2.4*	*1.1 – 6.0*		*1.2 – 7.0*
Number of compounds with no significant differences in peak area.		40			48			35	

Samples were profiled in negative nESI and positive nESI modes by nUHPLC-nESI-TOFMS. Mean peak areas of each analyte were calculated using 3 analytical replicates and significant differences in peak area between column types were determined using ANOVA (p<0.05).

Table 2.3: Comparison of mean peak area, retention time, percentage of organic solvent and column resolution between Acquity nUHPLC column types

	Column	PR (+)	T2 (+)	AN (+)	P (+)	TBEP (+)	T2G (-)	TXB2 (-)	TCA (-)	IBP (-)	AA (-)
Mean peak area (% CV)	HSS T3 100 Å	175200 (1)[a]	62899 (1)[a]	57489 (1)[b]	61870 (1)[b]	128058 (2)[b]	6369 (9)	924 (7)	8318 (6)[a]	531 (7)[a]	988 (3)[b]
	C18 130 Å	101830 (1)[c]	53242 (5)[b]	55844 (8)[b]	64218 (12)[b]	216234 (4)[a]	5551 (19)	703 (6)	5710 (2)[c]	338 (3)	168 (16)[c]
	C18 300 Å	114207 (4)[b]	33285 (3)[c]	22427 (1)[c]	23257 (1)[c]	35369 (1)[c]	7177 (10)[a]	758 (11)	7138 (13)[b]	362 (7)	4956 (2)[a]
Retention time (% CV)	HSS T3 100 Å	10.0 (0)	15.8 (0.02)	17.6 (0.02)	23.9 (0.02)	27.3 (0.02)	11.5 (0.2)	12.5 (0.05)	13.1 (0.06)	22.01 (0.08)	32.4 (0.06)
	C18 130 Å	9.7 (0.1)	13.8 (0.3)	15.1 (0.2)	20.8 (0.3)	25.8 (0.2)	10.7 (0.06)	11.8 (0.01)	11.5 (0.01)	20.3 (0.04)	31.4 (0.1)
	C18 300 Å	9.8 (0.4)	12.8 (0.05)	14.0 (0.3)	19.4 (0.5)	24.1 (0.3)	10.2 (0.2)	11.2 (0.3)	10.7 (0.3)	18.7 (0.4)	30.5 (0.1)
% organic solvent at elution	HSS T3 100 Å	38.4	46.5	49.0	74.8	89.1	40.5	41.9	42.7	66.8	100
	C18 130 Å	38.0	43.7	45.5	61.8	82.8	39.4	40.9	40.5	59.7	100
	C18 300 Å	38.1	42.3	44.0	55.9	75.6	38.7	40.1	39.4	52.9	100
Column resolution (% CV)	HSS T3 100 Å	N/A	53 (6)	64 (5)	101 (4)	156 (1)	N/A	76 (3)	21 (4)	101 (3)	321 (1)
	C18 130 Å	N/A	36 (3)	47 (4)	85 (8)	118 (4)	N/A	46 (1)	7 (7)	77 (10)	183 (3)
	C18 300 Å	N/A	24 (10)	33 (20)	79 (8)	133 (13)	N/A	66 (12)	9 (19)	102 (5)	382 (3)

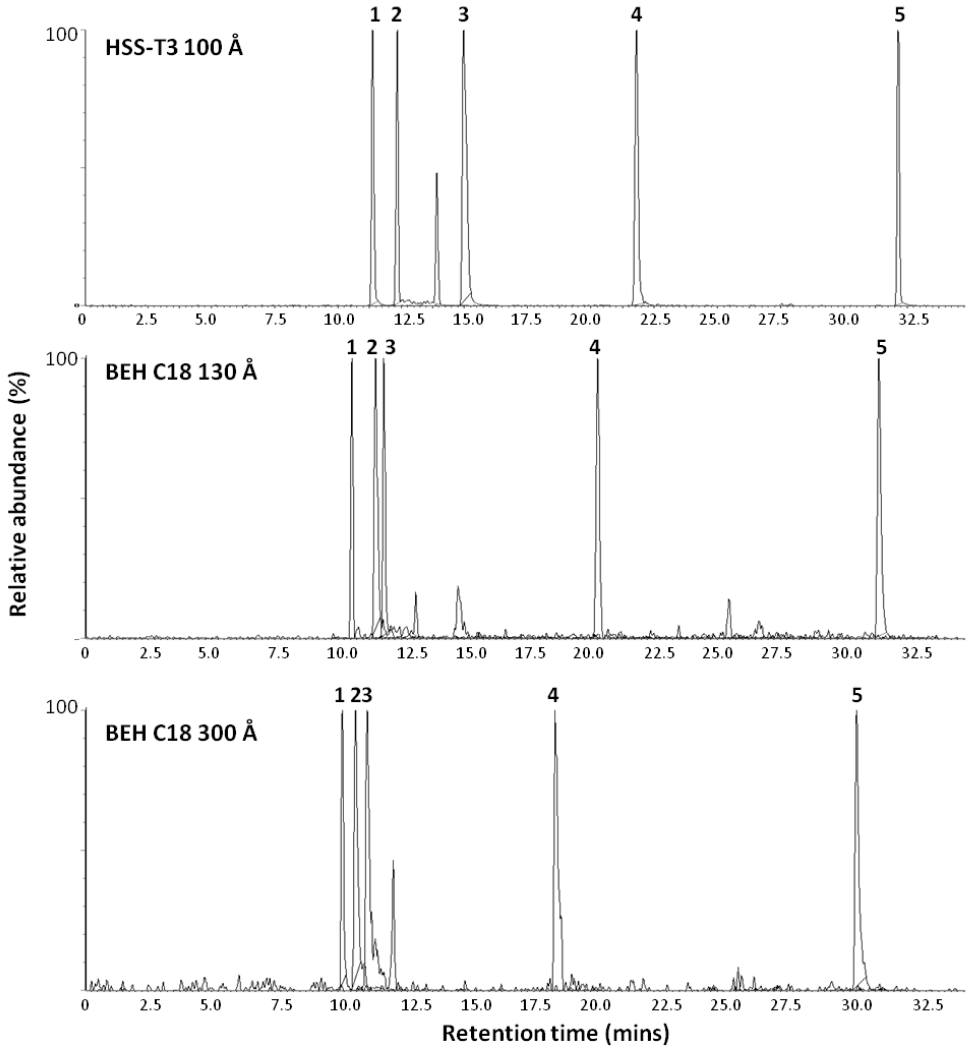
Samples were profiled in negative nESI (-) and positive nESI (+) modes by nUHPLC-nESI-TOFMS. CV: coefficient of variation, PR: propranolol, T2: testosterone, AN: androstenedione, P: progesterone, TBEP: tris(2-butoxyethyl)phosphate, T2G: testosterone glucuronide, TXB2: thromboxane B2, TCA: taurocholic acid, IBP: ibuprofen, AA: arachidonic acid. Means were calculated from three replicate injections. [a] = significantly higher response compared with the two other columns, [b] = response significantly higher than [c] (p<0.05). The benefits provided by the HSS T3 100 Å stem from its smaller pore size relative to the 130 Å and 300 Å BEH C18 columns and the reduced density of the trifunctional C18 chemistry of the T3 stationary phase compared to the C18 stationary phase.



1=unretained polar compounds, 2=metropolol, 3=venlafaxine, 4=propranolol, 5=carbamazepine, 6=testosterone, 7=androstenedione, 8=sphingosine, 9=diazinon, and 10=tris(2-butoxyethyl)phosphate. TIC=total ion count.

Figure 2.1: Base peak intensity (BPI) chromatograms of a standard mixture of compounds analysed by 3 different nUHPLC columns

100 pg of each compound was injected onto each column and the mixture analysed by nUHPLC-nESI-TOFMS in +nESI mode. A standard mixture of 78 compounds was analysed, of which 49 were detected by positive nESI. The BPI peaks of representative standards are labelled to illustrate changes in retention on the different column phases.



1=testosterone glucuronide, 2=thromboxane B2, 3= taurocholic acid, 4=ibuprofen, 5=arachidonic acid

Figure 2.2: Overlaid selected ion chromatograms standards analysed by 3 different nUHPLC columns

100 pg of each compound was injected onto each column and the mixture analysed by nUHPLC-nESI-TOFMS in negative nESI mode. A standard mixture of 78 compounds was analysed, of which 29 were detected by -ESI. Due to the signal strength not being high enough to give a BPI response the selected ion chromatograms of those standards highlighted in Table 2.3 are shown. These were selected ion chromatograms were achieved using a 5 ppm mass window of the accurate *m/z* provided in Table 2.1. The HSS-T3 100 Å column displays superior retention and separation characteristics of the standards compared to both BEH C18 columns.

2.4.2 Comparison of nUHPLC-nESI-TOFMS to a conventional UHPLC-TOFMS

The IDLs of the 78 standards injected on the nUHPLC-nESI-TOFMS were compared to that of a conventional UHPLC-ESI-TOFMS platform commonly used for metabolomic analysis (Al-Salhi et al., 2012). The amount of each of the 78 standards injected on both platforms ranged from 0.05 to 2000 pg. Results showed that the nanoflow-nanospray platform improved the detection of all 78 compounds by between 2-2000 times compared to the conventional UHPLC-ESI-TOFMS platform (Table 2.4). Furthermore, compounds such as estrone, 17β-estradiol, ethinylestradiol or bisphenol A, could not be detected at the highest concentrations used by conventional UHPLC-ESI-TOFMS platform, whereas the IDL for these compounds was between 1 and 25 pg on column for the nano platform. The IDL of glucocorticoids and androgenic compounds were between 25 to 50 pg on UHPLC-ESI-TOFMS, while the IDLs for these compounds were generally 0.05 pg on the nano platform. Important signalling molecules such as eicosanoids, which are present in biological fluids at nM levels, showed increases of 10-100 fold in IDL on the nano system with IDLs of < 1 pg. Additionally, the IDLs of pharmaceuticals such as non-steroidal anti-inflammatory drugs (NSAIDs), selective serotonin reuptake inhibitors (SSRIs), statins and β-blockers were improved by 2.5 to 100 fold on nUHPLC-nESI-TOFMS. Therefore, the use of this nanoflow-nanospray platform conferred a significant advantage in terms of instrumental detection for all chemical classes. Nevertheless, it should be noted that two different mass spectrometers were used for this comparison, but it can be assumed that improvements in IDLs were for the most part due to the improved desolvation, ionization and ion-transfer efficiency of the nUHPLC-nESI. Indeed, we observed that IDL of two compounds present in the mixture (i.e., carbamazepine and paracetomol) were 50 and 250 times greater using the same MS (i.e., the Xevo G2 TOFMS) but with conventional UHPLC-ESI platform (Vergeynst et al., 2013) showing the advantage of using a nUHPLC-nESI. Improvements made were compound specific and were observed in particular for important signalling compounds such as unconjugated steroids which are usually poorly ionized by a conventional UHPLC-ESI platform.

Table 2.4: Instrumental limits of detection (IDL) of a mixture of 78 compounds injected on a nanoflow-nanospray platform (nUHPLC-nESI-TOFMS) and a conventional UHPLC-ESI-TOFMS

Compound	IDL (pg)		Compound	IDL (pg)	
	ESI	nESI		ESI	nESI
Neurotransmitters			**NSAIDs**		
Epinephrine	200	25	Diclofenac	10	0.5
Dopamine	>2000	25	Ibuprofen	400	0.5
Serotonin	200	0.5	Naproxen	200	25
Melatonin	25	0.5	Ketoprofen	5	0.05
5-Hydroxyindoleacetic acid	200	10	Mefenamic acid	5	0.05
Nucleotides			Sulfasalazine	25	10
Cytidine	200	50	Indomethacin	25	10
Cytosine	25	10	Paracetomol	100	0.5
Amino acids and metabolites			**Lipid lowering agents**		
Histidine	2000	10	Clofibric acid	5	0.5
Tryptophan	600	50	Gemfibrozil	5	0.5
Creatinine	400	10	Bezafibrate	5	0.05
Vitamins			Pravastatin	5	0.5
Nicotinic acid	600	1.0	Simvastatin	25	10
Retinoic acid	>2000	10	**Heart disease drugs**		
Folic acid	2000	25	Dipyridamole	1	0.05
Steroids			Clopidogrel	5	0.05
Cortisol	50	0.05	Warfarin	0.5	0.05
Cortisone	50	0.05	**SSRI and antipsychotics**		
Corticosterone	25	0.05	Carbamazepine	5	0.05
11-ketotestosterone	10	0.05	Fluoxetine	25	1.0
21-Hydroxypregnenolone	100	1	Paroxetine	5	0.05
17α-Hydroxyprogesterone	50	0.05	Sertraline	50	0.5
Androstenedione	50	0.05	Venlafaxine	1	0.05
Progesterone	50	0.05	Clozapine	5	0.5
Pregnenolone	50	1	**β-Blockers**		
Testosterone	25	0.05	Metoprolol	1	0.05
Testosterone glucuronide	5	0.05	Propranolol	5	0.05
Estrone	>2000	10	Atenolol	25	0.5
Estrone glucuronide	5	0.5	**Pesticides-fungicides-antibacterial**		
Estrone sulfate	1	0.5	Diazinon	5	0.05
Estradiol	>2000	25	Piperonyl butoxide	1000	10
Estradiol glucuronide	5	0.5	Terbutryn	1	0.05
Eicosanoids			Atrazine	5	0.05
Prostaglandin E2	5	0.5	Miconazole	50	0.05
Prostaglandin B2	1	0.05	Propiconazole	5	0.05
Thromboxane B2	5	0.05	Triclosan	50	0.5

Table 2.4 continued					
Fatty acid metabolites			Chlorophene	5	0.05
Arachidonic acid	1000	0.5	**Endocrine disruptors/metabolites**		
Sphingosine	100	0.05	Bisphenol A	>2000	10
Bile acids			Butylparaben	5	0.5
Taurocholic acid	5	0.05	Ethinylestradiol	>2000	100
Cholic acid	5	0.5	Ethinylestradiol glucuronide	5	0.5
Taurodeoxycholic acid	5	0.05	Ethinylestradiol sulfate	5	0.5
Food additives			**Flame retardants**		
Caffeine	100	1	Triphenyl phosphate	5	0.05
			Tris(2-butoxyethyl)phosphate	25	0.05

The same standard mixture was injected on both platforms, the amount injected on column was between 0.05 pg and 2000 pg.

2.4.3 Repeatability and reproducibility of the nUHPLC-nESI-TOFMS method

The repeatability and reproducibility of the nUHPLC-nESI-TOFMS method was assessed using the composite urine samples spiked with deuterated IS and injected 10 times in both ionization modes, on 3 consecutive days. The coefficient of variation (CV) was calculated for the mean peak area and retention time of the 6 IS (see Table 2.5). The intra-day repeatability for peak area was between 2 and 19% for the urine matrix (analysed in both ionization modes). Inter-day reproducibility of the peak areas of the IS was also less than 20% which is the maximum variability recommended for metabolomics analyses using conventional LC-MS platforms such as UHPLC-ESI-TOFMS platforms (Benton et al., 2012). Retention time variability was less than CV 0.3%, and lower than a value of 2% recommended for conventional LC-MS platforms (Theodoridis et al., 2012). Furthermore, error in mass accuracy of the IS analysed by the nano platform across the 3 days was less than 1 ppm in positive nESI mode and less than 2 ppm in negative ESI mode.

Further analyses of the urine (xeno)metabolome itself were performed based on the repeatability of the response of the MS signals common to the analytical replicates. These analyses revealed that between 82%-90% of signals in urine returned CV values of less than 30% for signal response (Table 2.5). These values compare favourably with those recommended for metabolomics analyses using conventional LC-MS platforms, where maximum CV values of 30% were reported for the responses of the majority of MS features detected in urine or plasma samples (Benton et al., 2012, Want et al., 2010, Spagou et al., 2011). In this study the inter-day reproducibility of the (xeno)metabolome datasets was generally lower; with 73% (positive nESI) and 81% (negative nESI) of all signal responses gave a CV of < 30%. To date, no recommendations for inter-day variation of the metabolome have been reported due to confounding factors on the variability of analytical response such as column aging, source fouling and the effect of cleaning (Benton et al., 2012). However, these

results indicate that this nanoflow-nanospray platform offers a high degree of stability which is necessary for large scale studies of the (xeno)metabolome.

Table 2.5: Intra and inter-day repeatability based on measurements of the response of internal standards spiked and variations of the normalised abundance of the features detected in all replicates in urine samples analysed by nUHPLC-nESI-TOFMS

		Urine			
Sample		D1	D2	D3	D1/2/3
Venlafaxine-d6 (+ESI)	Mean area	83699	76457	75709	80078
	CV (%)	2	5	2	6
	Mean Rt	9.53	9.49	9.46	9.51
	CV (%)	0.09	0.08	0.12	0.3
Carbamazepine-d10 (+ESI)	Mean area	251578	217009	179895	234293
	CV (%)	5	7	6	10
	Mean Rt	11.93	11.91	11.92	11.92
	CV (%)	0.07	0.05	0.06	0.11
Progesterone-d9 (+ESI)	Mean area	110318	98459	119646	104389
	CV (%)	2	10	8	8
	Mean Rt	23.66	23.65	23.66	23.66
	CV (%)	0.05	0.05	0.04	0.04
Estradiol sulfate-d4 (-ESI)	Mean area	73243	62479	59983	77742
	CV (%)	10	7	17	18
	Mean Rt	11.06	11.06	11.05	11.06
	CV (%)	0.07	0.05	0.11	0.09
Prostaglandin E2-d4 (-ESI)	Mean area	10519	8619	7354	8739
	CV (%)	19	5	18	20
	Mean Rt	14.06	14.06	14.07	14.06
	CV (%)	0.1	0.06	0.06	0.08
Diclofenac-d4 (-ESI)	Mean area	9306	9875	10472	9974
	CV (%)	13	5	8	7
	Mean Rt	21.69	21.7	21.72	21.7
	CV (%)	0.07	0.02	0.06	0.07
MS signals common to all sample replicates (+ESI)	Percentage of MS signals with CV < 30%	88	90	82	73
	Total number of signals	457	482	436	348
MS signals common to all sample replicates (–ESI).	Percentage of MS signals with CV < 30%	83	82	82	81
	Total number of signals	590	605	633	444

Analytical replicates (n=10) were analysed each day for 3 days in (+) positive nESI or (–) negative nESI modes. IS repeatability and reproducibility were assessed using the CV of mean peak area and retention time. (Xeno)metabolome variations were measured by calculating the CV of the response of MS signals common to the analytical replicates. The number of signals with a % CV <30 are expressed as a percentage of total.

2.5 Conclusion

Nanoflow LC systems such as nUHPLC-nESI-TOFMS are routinely used for proteomics applications but to date are not generally used for metabolomics because of the high degree of repeatability necessary for processing and analysis of the datasets. For these small molecule profiling studies, the MS features are generally binned into Rt x m/z values prior to multivariate analyses. In past work, variability in nanoLC flow rates and spray emission into the nanoESI source have been too high for metabolomics applications. However, using a nUHPLC

system and a flow rate of 700 nL/min, this study has shown repeatability and reproducibility values comparable to conventional UHPLC-ESI-MS systems. In addition, compared to a conventional ESI platform that is typically used for metabolomics, there were between 2 and 2000 fold increases in IDL with the nano system, which could enable wider coverage of the (xeno)metabolome including detection of low abundant or poorly ionized metabolites. The choice of nano column phase and porosity can also influence analyte signal and retention, although improvements in peak areas using low porosity 100 Å phases can be compound dependent. This study reveals that use of a nanospray platform could be invaluable to support the rapidly growing field of metabolomics analyses, where the ability to reliably detect low abundance metabolites and xenobiotics in biofluids may pave the way for discovery of new biomarkers and pathways of disease.

Chapter 3: Development of solid phase extraction and nanoflow liquid chromatography-nanoelectrospray ionization mass spectrometry for improved global urine metabolomics

3.1 Abstract

Global urine metabolomics is a rapidly expanding field with the potential to discover biomarkers of disease and exposure. To date most methods focus on rapid sample preparation, using neat or diluted urine, together with high throughput analyses, and are poorly suited for detection of low abundance metabolites present in urine samples. In this study, novel methods have been developed to analyse urine by nUHPLC-nESI-TOFMS after preconcentration by solid phase extraction (SPE), thus enabling significant improvements in analytical sensitivity and coverage of the urinary metabolome. In initial work, urine samples were extracted by both anion and cation exchange mixed mode polymeric SPE cartridges and qualitatively compared with those using conventional sample preparations using UHPLC-ESI-TOFMS analyses. Compared with neat or diluted urine samples, SPE concentration of urine resulted in detection of additional metabolites including bile acids, lipids, pharmaceuticals and markers of lifestyle, with little loss of other components of the metabolome. Analyses of SPE preparations by nUHPLC-nESI-TOFMS revealed excellent retention time repeatability with <1% coefficient of variation (CV) for 96% of analysed peaks. The repeatability of the MS response was <30% CV for >79% of MS features in both negative and positive nESI modes. Compared with UHPLC-ESI-TOFMS, analysis by the nano-platform enabled detection of signalling molecules important in disease processes including sex steroids, glucocorticoids, eicosanoids and neurotransmitter metabolites. The significant improvement in sensitivity arising from use of nUHPLC-nESI-TOFMS analyses of SPE-concentrated samples represents a step change in coverage of the urinary metabolome thereby increasing the potential for biomarker discovery.

3.2 Introduction

The aim of global metabolomics is to analyse all metabolites in a given biological system, termed the metabolome (Dunn et al., 2011, Dettmer et al., 2007). This technique is generally utilized to investigate metabolic variations between metabolomes in case study and control groups (Gika et al., 2014). As such, metabolomics has a wide range of applications including biomarker discovery of disease (Mamas et al., 2011), toxicant exposure (Al-Salhi et al., 2012), assessment of nutritional status (Llorach et al., 2012) and uses in drug discovery (Kell, 2006, Lindon et al., 2007). To date >3000 urinary metabolites have been characterised, a number that is expected to grow as more sensitive analytical techniques are developed (Bouatra et al., 2013). These metabolites are typically <1000 Da, cover a concentration range

of 11 orders of magnitude and include endogenous metabolites and those derived from exogenous sources such as drugs, contaminants or diet (Bouatra et al., 2013, Johnson et al., 2012a, Want et al., 2007). Global analysis of this complex mixture of metabolites in urine holds great promise for understanding the impacts of lifestyle, environmental stressors and disease on human health, and for identifying potential biomarkers of environmental exposure in a non-invasive manner (Pujos-Guillot et al., 2013, Zacharias et al., 2013, Cernei et al., 2013).[12]

Of the different analytical chemistry techniques that have been used to profile the urinary metabolome, the use of high resolution LC-MS is widely recognised to provide the most comprehensive analysis allowing detection of non-volatile and semi-volatile polar and apolar compounds (Gika et al., 2014, Theodoridis et al., 2012). However, to date sample preparation for LC-MS analyses of urine has received little attention (Theodoridis et al., 2012, Alvarez-Sanchez et al., 2010a, Alvarez-Sanchez et al., 2010b, Fernández-Peralbo and Luque de Castro, 2012) and many methods focus on rapid sample preparation and high throughput analysis instead of increasing the sensitivity for detection of important low abundance metabolites (Theodoridis et al., 2012). The sample preparation methods currently most utilised in urine metabolomic studies comprise just dilution of urine samples and these methods have been described in Chapter 1 Section 1.5.2.

To detect low abundance metabolites, further sample clean up and pre-concentration is required, and could be achieved using solid phase extraction (SPE) (Alvarez-Sanchez et al., 2010a, Alvarez-Sanchez et al., 2010b, Fernández-Peralbo and Luque de Castro, 2012). Typically SPE is a sample preparation method reserved for targeted analysis of selected compounds in urine and in global metabolomics studies there are concerns that use of SPE result in losses of many metabolites during sample work up (Fernández-Peralbo and Luque de Castro, 2012). This is a particular problem with urine which contains complex mixture of compounds from the polar amino acids and anionic conjugated metabolites, through to apolar lipids, steroids, bile acids and xenobiotics. To counter this, a method was developed in this study utilising one of each anion and cation exchange mixed mode polymeric SPE cartridges to sequester ionic, neutral and apolar metabolites and remove ion suppressing salts and proteins. Total evaporation of the SPE eluent and reconstitution for LC-MS analysis allows for >10 fold pre-concentration of sample extracts which may lead to improved detection of low abundance metabolites using non-targeted LC-MS techniques (Fernández-Peralbo and Luque de Castro, 2012).

The study from Chapter 2 has demonstrated that nUHPLC-nESI-TOFMs enables highly sensitive analysis compared to conventional LC-ESI-TOFMS and maintains a very high degree of

repeatability and reproducibility for peak area and retention times. However, the benefit of this technique is yet to be shown in a metabolomic analysis.

3.2.1 Study aims

The aim of this study was to develop methods to increase coverage of the urinary metabolome including the detection of low abundant signalling metabolites normally present in urine. To determine the most suitable sample preparation technique, the metabolomic profiles of neat and diluted urine preparations were qualitatively compared with samples extracted by SPE using conventional UHPLC-ESI-TOFMS. The metabolomic profiles of SPE extracts were subsequently analysed using conventional UHPLC-ESI-TOFMS and nUHPLC-nESI-TOFMS to investigate improved metabolome coverage due to use of the nanospray platform.

3.3 Materials and Methods

3.3.1 Chemicals

All HPLC grade solvents were supplied by Rathburn Chemicals Ltd (Walkerburn, UK) and UHPLC grade solvents from Fisher Scientific UK (Loughborough, UK). Mixed mode polymeric anion (Strata X-AW) and cation exchange (Strata X-C) SPE cartridges (60 mg/3 mL) were purchased from Phenomenex Ltd (Macclesfield, U.K). Deuterated progesterone-2,2,4,6,6,17α,21,21-d9 (CDN Isotopes, Quebec, Canada) and 17β-estradiol 2,4,16,16-d4 sodium, 3-sulfate (Cambridge Isotope Laboratories Inc, MA, US) were used as internal standards (IS). All other chemicals were purchased from Sigma-Aldrich Ltd (Dorset, UK).

3.3.2 Sample collection

Urine samples were collected from two self-reported healthy volunteers (1 male and 1 female) aged 23 and 24. The volunteers collected first morning voids, mid-flow, over a 10 day period. Samples were stored immediately in 10% methanol and transported to the laboratory in a cool box and stored at -80 ^{o}C. Both subjects provided written, informed consent and the study was approved by the University of Sussex Life Sciences and Psychology Ethics Committee.

3.3.3 Sample preparation

Urine samples were defrosted and vortex mixed. 10 mL aliquots from each first morning void were combined to produce a 100 mL pooled male and a pooled female sample of neat urine. Three diluted samples were prepared by the addition of HPLC grade water to aliquots of the pooled samples to give 25%, 50% and 75% concentrations of the neat urine, equivalent to 1: 3, 1: 1 and 3: 1 (urine: water). Urine samples (200 µL) were spiked with the

two stable isotope internal standard (IS) (200 ng) and centrifuged at 13000 rpm for 10 minutes. The resulting supernatant was passed through a Strata Impact Protein Precipitation 2 mL Filter Plate (Phenomenex Ltd, Macclesfield, U.K) to remove particulates.

For SPE of samples, aliquots (2 mL) from each of the pooled neat urine samples were spiked with IS to give 0.5 ng/µL final concentration and the pH adjusted with formic acid (FA) to pH 2. Strata X-AW and Strata X-C cartridges were conditioned with 2 mL methanol and 2 mL HPLC water. The Strata X-AW was loaded with 2 mL of the acidified urine; the load that passed through was collected and directly loaded onto the Strata X-C cartridge. Both cartridges were washed with 2 mL water, each cartridge was eluted separately with 2 mL 5% ammonium hydroxide in methanol followed by 2 mL ethyl acetate. The two solvent elutions from each cartridge were combined, the solvents removed under vacuum and the extracts reconstituted in 200 µL 80:20 water: methanol, giving a 10 fold concentration of the original 2 mL sample. For some studies, aliquots of extracts from both the X-AW and X-C SPE cartridges were combined to allow analysis of the total SPE preparation.

Five replicates of neat, diluted and SPE samples were prepared from the pooled male and female urine samples using the methods described. Work up blanks were produced and consisted of 90% water 10% methanol. All samples were frozen at -80 ^{o}C prior to LC-MS analysis.

3.3.4 UHPLC-ESI-TOFMS analysis

Samples (5 µL) were analysed in both ionization modes in a randomised block design using a Waters Acquity UHPLC coupled with a Waters Ultima TOFMS system (Waters Ltd, Manchester, UK). The MS system was conditioned using repeated analyses of a quality control (QC) sample prior to analysis of the test replicates. Separation was carried out on an Acquity BEH C18 column (2.1 x 100 mm, 1.7 µm) at 25^{o}C, with a C18 vanguard (2.1 x 50 mm) column. Mobile phase A consisted of water (0.2% FA) and B acetonitrile (0.2% FA). The gradient was 0-1 min 5% B, 1-5 min 15% B, 5-18 min 40% B, 18-25 min 100% B, 25-26 min 100% B at a flow rate of 0.2 mL/min. Prior to each injection, the column was equilibrated in initial conditions for 5 minutes. The mass spectrometer was tuned to a mass resolution of 9000 and mass spectra data were collected between 100 to 1000 *m/z*. A lockmass solution was infused during analysis, and comprised sulfadimethoxine in 1:1 v/v methanol: water (plus 0.1% FA for acquisition in positive ESI mode). The source temperature and desolvation temperatures were 100 ^{o}C and 250 ^{o}C respectively, with a desolvation gas flow of 400 L/h nitrogen.

3.3.5 NanoflowUHPLC-nanoESI-TOFMS

SPE extracts were analysed on a Waters nanoAcquity UHPLC, and 0.5 µL aliquots of extract were injected on a Waters UHPLC HSS-T3 nanoAcquity column (100 Å, 1.8 µm, 100 µm x 100 mm) maintained at 25 ^{o}C. Mobile phase A was water (0.01% FA) and B acetonitrile (0.01% FA) at a flow rate of 700 nL/min. Initial conditions were 10% B with a gradient of 0-4 min to 30% B, 4-18 min 50% B, 18-28 mins 100% B. The column was maintained at 100% B for 10 minutes to and then re-equilibrated in initial conditions for 15 minutes before the next injection. The sample needle was washed in 1.5 mL water and 1.5 mL acetonitrile to eliminate carry over in subsequent injections. SPE samples were analysed after conditioning the MS system with QC samples. The Waters Xevo G2 with nanospray source was tuned to a mass resolution of 20,000 with a mass spectra range of 50-1000 *m/z*. A capillary voltage of 3.5 kV was used in both ESI modes, with collision energy of 10 eV and a flow of 61 L/hr for the cone and 300 L/hr for desolvation nitrogen gas. A leucine enkephalin lockspray standard in 1:1 methanol:water (2 ng/µL) was infused at 700 nL/min in both ionization modes. Mass fragmentation information was obtained by further analysis of selected samples in MS^{e} mode, this mode allows rapid voltage ramping to achieve low and high fragmentation data concurrently, here a scan at 10 eV was followed by a scan with a ramp of 20 to 30 eV.

3.3.6 Nanospray emitter construction

Nanospray emitters were produced in-house and a non conducting sheath was constructed from polyetheretherketone (PEEK) tubing (100 µm I.D to a length of 6 cm (Kinesis, St Neots, UK). The emitter, a fused silica tube, 10 µm I.D. and 90 µm external diameter (Waters, Manchester, UK), was cut to 8 cm length and passed through the PEEK tubing. Subsequently, all emitters were sonicated in HPLC grade MeOH for 10 min prior to use.

3.3.7 Quality control (QC) and methodological reproducibility

QC samples, comprising a composite of the samples under analysis, were injected prior to and during each analytical batch to monitor instrumental drift in response and retention time data. Retention time repeatability of the nano platform was determined as the % coefficient of variation (CV) of the mean retention time. This was assessed for the 100 most abundant metabolites (50 +nESI, 50 -nESI), which eluted across the chromatogram of the 10 SPE QC samples (Theodoridis et al., 2012). The 100 most abundant peaks were extracted from the 10 QC replicates using Waters Quanlynx software. In addition the peak area variability of the QC samples was determined using methodology proposed by Want et al. For this, all peaks present in at least 80% of the QC samples were integrated and the % CV of mean peak areas calculated. The results were compared with the recommended threshold for reliable

metabolomic data whereby the % CV for the mean peak area should be <30% for >70% of peaks (Want et al., 2010). The repeatability of the SPE methodology was similarly determined for metabolome peak areas present in 80% of the samples after UHPLC-ESI-TOFMS analysis for each of the 5 male and 5 female SPE replicates.

3.3.8 Data processing

The MS datasets were deisotoped, mass spectral peaks deconvoluted, aligned, and the datasets binned and normalized to the total spectral area for each sample using Waters MarkerLynx software. The datasets, comprising Rt x *m/z* bins, were exported to Simca v12.0 software (Umetrics Ltd, Crewe, UK) for multivariate analyses. All data were log transformed and Pareto scaled prior to principal components analysis (PCA) to identify the effect of different sample preparation treatments. Further modelling using orthogonal partial least-squares discriminant analysis (OPLS-DA) was used to investigate metabolite differences between any two treatment groups. Discriminatory metabolites (loading variables) between treatment groups were detected using an S-plot of the OPLS-DA model which is a plot of reliability (correlation) of the loading variables versus their covariance (contribution to the model) (Wiklund et al., 2008). Metabolite identities were determined from their accurate mass, isotopic fit, and comparison of fragmentation data with authentic standards or with Metlin (Tautenhahn et al., 2012a), Human Metabolome Database (Wishart et al., 2007), Human Urine Metabolome Database (Bouatra et al., 2013) and MycompoundID (Li et al., 2013a) databases.

3.4 Results and discussion

3.4.1 Effect of sample preparation methodology

The effect of sample dilution and SPE on the MS profiles of the urine metabolome was investigated. Replicate analytical samples prepared from the pooled urine samples from both subjects were profiled by UHPLC-TOFMS, and the number of markers (MS features) calculated from the binned MS data. The mean number of markers detected in each of the two neat urine samples was between 4400 and 5400, and dilutions to ≤50% urine significantly decreased marker numbers (ANOVA $p<0.01$) (Figure 3.1). Compared to neat urine samples, analysis of urine extracted by SPE (a combined extract from both X-AW and X-C cartridges) resulted in a significant (ANOVA $p<0.01$) 1.7 fold increase in the number of markers detected in either + or - ESI mode.

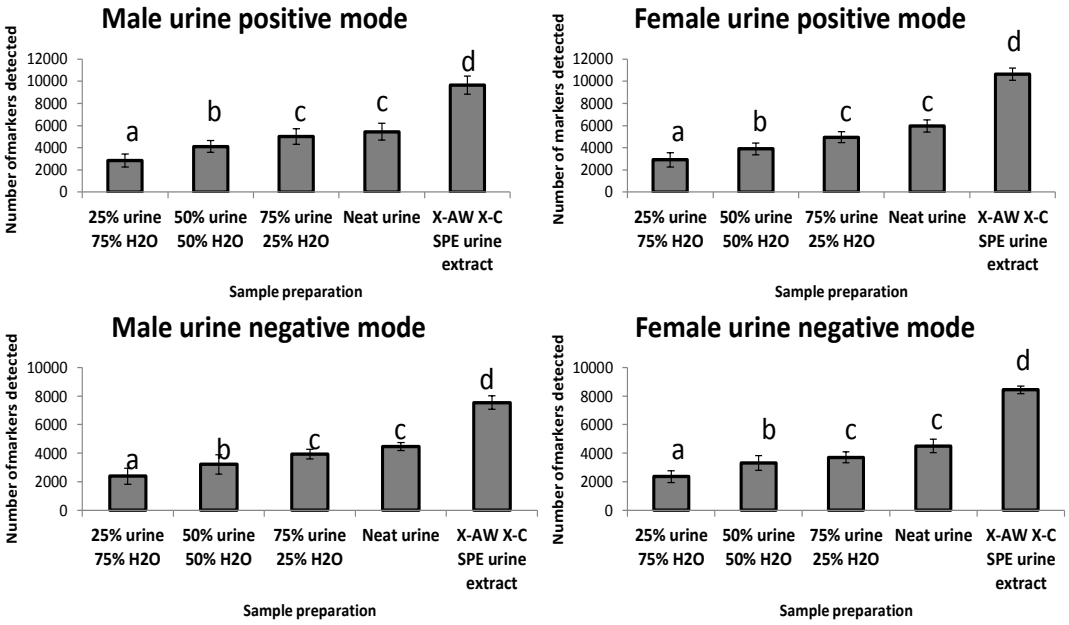


Figure 3.1: Numbers of markers detected using different urine sample preparation methods

Number of markers are given as mean ±SD (n=5) after removal of workup blank signals which were 305 ±47 (+ESI) and 140±9 (-ESI). Letters above histogram indicate significant differences after ANOVA analyses and post hoc Tukey's test. Same letters indicate no significant difference ($p \leq 0.05$), different letters mark a significant difference compared to neat urine ($p \leq 0.01$).

Qualitative differences in MS profiles between the sample preparations were examined by multivariate methods. PCA scores plots of the neat and diluted urine datasets revealed clear differences between the sample treatments (Figure 3.2 and 3.3). The first component discriminated between the two individuals providing the urine samples and the second component differentiated between the sample preparation methods. The close grouping of replicates from any one sample preparation method indicated that the 5 analytical replicates in the sample groups were very similar in composition to each other. This is further supported by the very close grouping of the QC samples, indicating, the MS profiling methodology was repeatable. The datasets from neat urine were compared with each of the diluted urine preparations in turn using OPLS-DA models (see Figure 3.4A for example). The markers that differentiated between the two sample preparations treatments were detected using an S-plot. Markers that were either unique to, or were significantly higher, in one of the sample preparation groups were detected at the extreme ends of the S-plot (Figure 3.4B). Examination of the S-plots revealed that there were 23 metabolites present in neat or 75% diluted male or female urine samples which were not detected after dilution to ≤50% urine. Metabolites that were unique to the neat urine and absent in samples diluted by 50% comprised some conjugated steroids, several bile acids, vitamins, and markers of lifestyle and diet such as cotinine and sucralose (Table 3.1). These have reported concentrations in urine of

generally < 1 µM/mM creatinine (Bouatra et al., 2013), and have likely been diluted to such a degree they can no longer be detected.

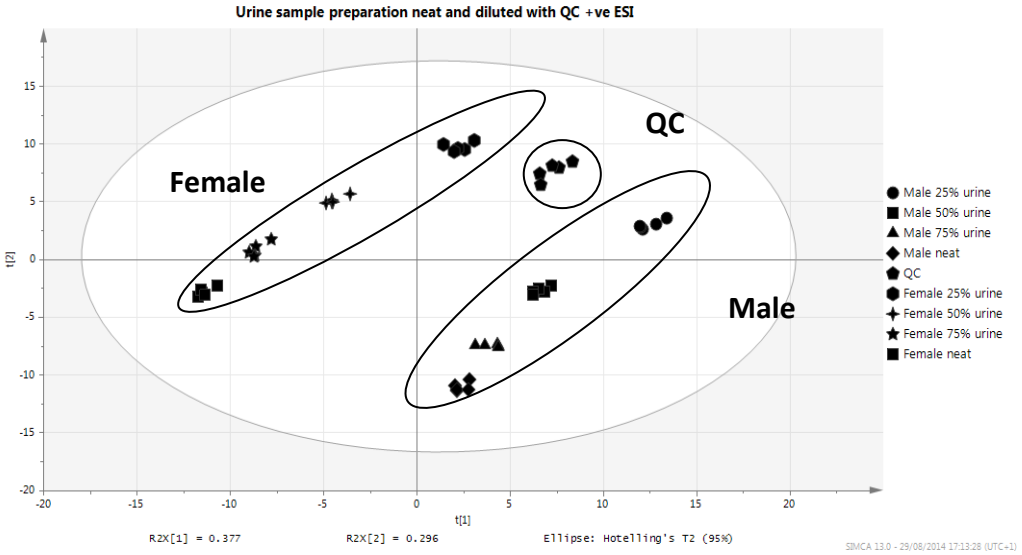


Figure 3.2: PCA scores plot of the metabolomic profiles from neat, diluted and QC urine sample preparations

Neat urine from either a male or female subject was diluted with water to give 25%, 50% and 75% urine concentrations. An aliquot (5 µL) of each of the 5 replicates was analysed in positive (+ve) ESI by UHPLC-ESI-TOFMS.

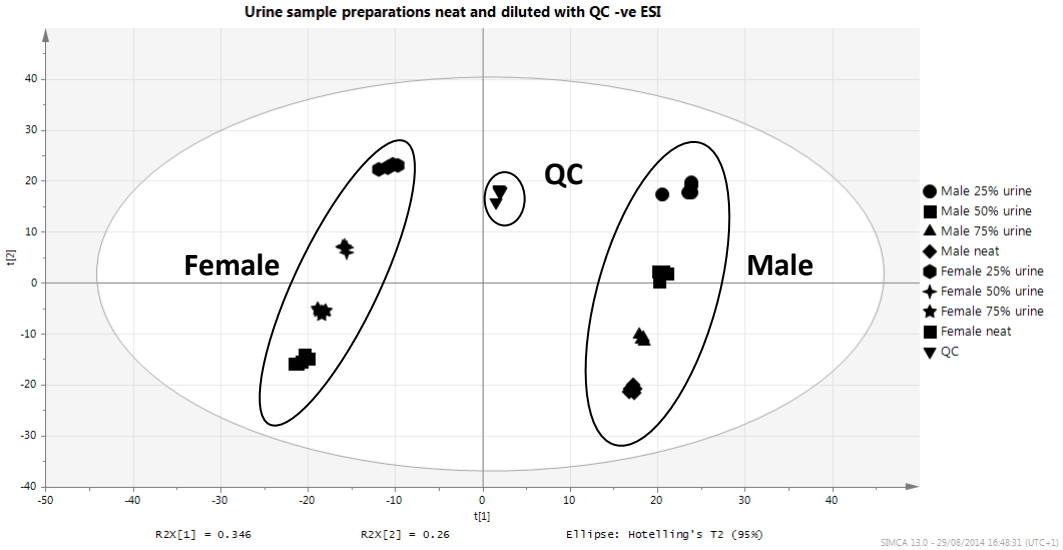


Figure 3.3: PCA scores plot of the metabolomic profiles from neat, diluted and QC urine sample preparations

Neat urine from either a male or female subject was diluted with water to give 25%, 50% and 75% urine concentrations. An aliquot (5 µL) of each of the 5 replicates was analysed in negative (-ve) ESI by UHPLC-ESI-TOFMS.

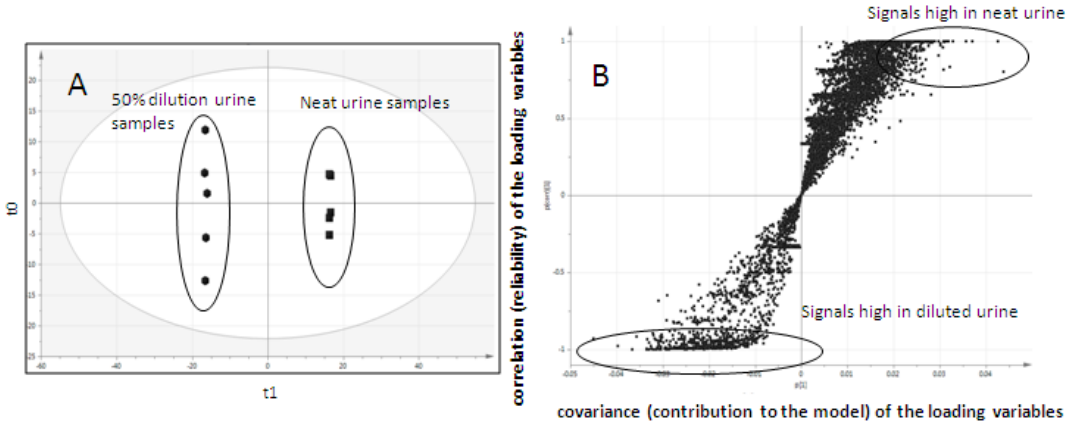


Figure 3.4: A) OPLS-DA scores plot and B) S-plot analysis of metabolomics profiles from neat and diluted (50%) urine sample preparations

Data shown for 5 replicates of neat and diluted (50%) urine preparations from a male urine sample analysed in negative ESI by UHPLC-ESI-TOFMS. A) OPLS-DA model, B) S-plot analysis of loading variables (MS signals).

Table 3.1: Metabolites unique to neat urine compared with samples diluted to 50% concentration

Observed m/z	Empirical formula	Theoretical mass	Q-TOF Fragments	Identity
Amino acids and metabolites				
M+H 156.0762	$C_6H_9N_3O_2$	156.0773	110.0685	Histidine
M-H 245.0930	$C_{13}H_{14}N_2O_3$	245.0926	203.0801, 74.0254	N-acetyltryptophan
M+H 195.0769	$C_9H_{10}N_2O_3$	195.0770	149.071, 120.045	Aminohippuric acid
Neurotransmitter/ metabolites				
M-H 182.0822	$C_9H_{13}NO_3$	182.0817	166.084	Normetanephrine
Vitamins				
M+H 377.1456	$C_{17}H_{20}N_4O_6$	377.1461	243.0867	Riboflavin (Vitamin B2)
M-H 297.1128	$C_{18}H_{18}O_4$	297.1127	nd	7C-aglycone[1] (a vitamin K metabolite)
Conjugated androgen				
M-H 479.2279	$C_{25}H_{36}O_9$	479.2281	175.0224, 113.0223	11-Oxo-androsterone glucuronide[1]
M-H 467.2641	$C_{25}H_{40}O_8$	467.2645	175.0239, 157.0145, 113.0241	Androstanediol glucuronide[1]
Glucocorticoids				
M-H 535.2170	$C_{27}H_{36}O_{11}$	535.2179	175.0241, 157.0138, 113.0242	Aldosterone glucuronide[1]
Bile acids				
M-H 407.2798	$C_{24}H_{40}O_5$	407.2797	343.2623, 289.2140, 251.2000	Cholic acid
M-H 567.3167	$C_{30}H_{48}O_{10}$	567.3169	175.0265, 113.0318	Deoxycholic acid glucuronide[1]
M-H 624.3390	$C_{32}H_{51}NO_{11}$	624.3384	175.02298, 113.229	Glycochenodeoxycholic acid glucuronide[1]
Diet compounds				
M-H 395.0063	$C_{12}H_{19}Cl_3O_8$	395.0067	359.0368	Sucralose
M-H 297.0979	$C_{14}H_{18}O_7$	297.0974	175.0241,157.0142, 113.0242	Phenylethanol glucuronide[1]
M-H 507.2234	$C_{26}H_{36}O_{10}$	507.2230	nd	Unidentified dietary compound[1]
Smoking related compounds				
M+H 177.1031	$C_{10}H_{12}N_2O$	177.1028	137.0027	Cotinine
M+H 179.1177	$C_{10}H_{14}N_2O$	179.1184	nd	Nicotine-1'-N-oxide[1]
Organic acids ≤10 carbons				
M-H 201.1127	$C_{10}H_{18}O_4$	201.1127	183.1018, 139.1119	Sebacic acid
M+H 291.1312	$C_{10}H_{18}N_4O_6$	291.1305	nd	Argininosuccinic acid[1]
M-H 187.0973	$C_9H_{16}O_4$	187.0970	125.097,97.066, 57.036	Azelaic acid

Table 3.1 continued				
M-H 266.0869	$C_9H_{17}NO_8$	266.0876	nd	Neuraminic acid[1]
M+H 146.0929	$C_5H_{11}N_3O_2$	146.0930	104.0701, 87.0450, 86.1601	Guanidinobutanoic acid
Organic acids >10 carbons				
M+H 449.2554	$C_{18}H_{36}N_6O_5S$	449.2546	nd	Glutathionylaminopropylcadaverine[1]

Metabolite identities confirmed with standards or comparison with fragmentation pattern in databases. [1]Putative identity based upon accurate mass measurement of molecular ion and a knowledge of dietary and pharmaceutical intake in addition to the more common metabolites present in urine to make an informed decision as to the potential identity of metabolites where genuine standards or database fragmentation details are not available. nd = metabolite signal too weak for detection of fragments.

There were many other markers that were detected in neat urine but not diluted preparations. These were identified as dimer, trimer or adduct ions of metabolites. These were artefacts of the MS analysis which were present in spectra in addition to the molecular ion of the metabolite which was detected in all the diluted preparations. The reduced level of dimers and trimers in the diluted samples is likely due to the reduced concentration of highly abundant compounds in the ESI source which may be less favourable to their formation. Analysis of the other end of the S-plot revealed markers which were at higher relative concentrations in the diluted urine samples compared to the neat urine. Examination of the datasets confirmed that they were also detected in neat samples too but at lower concentrations, possibly due to greater ion suppressive properties of neat urine preparations (Waybright et al., 2006).

A similar comparison between the MS profiles of neat urine and sequential extraction with 10 fold concentration of urine by the two mixed mode polymeric cation and anion exchange SPE phases, revealed markers unique to each preparation method. To assess the repeatability of the SPE methodology, the variability of the mean peak areas in the MS profiles of replicate SPE extracts was assessed and compared to that of replicates of neat samples via the method proposed by Want et al., (2010). Analysis of replicate neat urine samples by either positive or negative ESI modes was highly repeatable and 91-93% of all 3503-4484 peaks detected in 80% of the samples gave a mean peak area %CV of <30%. Analysis of replicate SPE samples revealed that between 78% and 87% of 6684-9139 peaks returned mean peak areas of <30% CV. This indicates that SPE introduces an additional 6-15% of the experimental variation compared with direct analyses of urine samples, though the variability still falls well within the acceptable range suggested for metabolomics analyses; i.e. >70% of peaks present in 80% of samples should have a variability of <30% CV of the mean peak area (Want et al., 2010).

Compared to neat urine, SPE concentration of samples from the male or female subjects resulted in detection of a further 24 metabolites. These included metabolites of pharmaceuticals, diet and lifestyle as well as non-conjugated bile acids, lipid and amino acid metabolites (see Table 3.2). Hence, a tenfold concentration and improved sample clean up by SPE increased the detection of many other low abundant compounds to a level that can be detected and thus monitored for metabolic variance. Bile acids in particular are reported to be present at levels of <0.01 µM/mM creatinine in human urine (Bouatra et al., 2013), and are known to be indicative of liver toxicity (Kumar et al., 2012).

Conversely, 12 metabolites were unique to neat urine and not detected in the SPE samples. These comprised polar metabolites that were poorly retained on the UHPLC column,

eluting between 1.0-3.8 mins. This indicates that they were too hydrophilic to be extracted by the SPE phases. For these compounds analyses of neat or diluted urine by HILIC methods would be more suitable (Gika et al., 2014). These polar compounds included sugar, purine and nucleoside metabolites which were either endogenously derived or products of the diet. Many of the additional markers uniquely detected in concentrated SPE samples were due to the formation of dimer, trimer or adduct ions supplementary to the molecular ion signal of the metabolite which was present in both neat and SPE preparations. These results indicate that SPE extraction of urine results in little qualitative loss from the metabolome, and enable the removal of salts and concentration of organics from samples necessary for analyses on a nanoflow/nanospray MS platform.

Table 3.2: Metabolites unique to either SPE or neat urine preparations

Experimental mass and ion species.	Putative Structure	Theoretical mass	Fragments	Identity
		Metabolites unique to SPE		
Amino acids and metabolites				
M-H 192.0664	$C_{10}H_{11}NO_3$	192.0661	148.0752, 91.0573	Methylhippuric acid
M+H 170.0925	$C_7H_{11}N_3O_2$	170.0930	126.1021, 109.0751, 96.0683	Methyl histidine
M+H 295.1292	$C_{14}H_{18}N_2O_5$	295.1294	nd	Glutamylphenylalanine[1]
Carnitines				
M+H 372.3108	$C_{21}H_{41}NO_4$	372.3114	85.0272	Tetradecanoylcarnitine
Neurotransmitter/ metabolites				
M+H 249.1231	$C_{13}H_{16}N_2O_3$	249.1239	190.0858	Hydroxymelatonin
Vitamins				
M-H 605.3331	$C_{33}H_{50}O_{10}$	605.3326	175.0238,157.0146, 113.0243	Dihydroxy-oxovitamin D3-glucuronide[1]
Bile acids and metabolites				
M-H 611.3795	$C_{33}H_{56}O_{10}$	611.3799	175.0242,157.0149, 113.0244	Cholestane-tetrol-glucuronide
M-H 391.2843	$C_{24}H_{40}O_4$	391.2848	nd	Chenodeoxycholic acid
M-H 514.2836	$C_{26}H_{45}NO_7S$	514.2838	124.0071	Taurocholic acid
M-H 391.2848	$C_{24}H_{40}O_4$	391.2848	327.271, 329.2836, 345.2793, 347.2956	Deoxycholic acid
M-H 455. 2478	$C_{24}H_{40}O_6S$	455.2478	nd	Sulfolithocholic acid[1]
M-H 512.2692	$C_{26}H_{43}NO_7S$	512.2682	nd	Unidentified bile acid[1]
Lipid metabolites				
M+H 288.2911	$C_{13}H_{37}NO_2$	288.2903	nd	C17 Sphinganine[1]
M+H 542.2999	$C_{24}H_{47}NO_{10}S$	542.2978	nd	Psychosine sulfate[1]

Table 3.2 continued				
M+H 358.2957	$C_{20}H_{39}NO_4$	358.2963	nd	Palmitoyl threonine[1]
Dietary compounds				
M-H 415.1243	$C_{18}H_{24}O_{11}$	415.1240	nd	Hydroxybenzaldehyde-xylosyl-glucoside[1]
M-H 212.0024	$C_8H_7NO_4S$	212.0018	132.0455	Indoxylsulfuric acid
M-H 258.9910	$C_9H_8O_7S$	258.9912	nd	Caffeic acid sulfate[1]
Smoking related compounds				
M+H 163.1235	$C_{10}H_{14}N_2$	163.1238	132.0785, 130.0785	Nicotine
M+H 163.1229	$C_{10}H_{14}N_2$	163.1235	146.0951	Anabasine
Free pharmaceuticals				
M+H 223.1328	$C_{13}H_{18}O_3$	223.1334	nd	Hydroxyibuprofen[1]
M+H 152.0704	$C_8H_9NO_2$	152.0712	nd	Acetaminophen
Conjugated pharmaceuticals				
M-H 285.0789	$C_{13}H_{18}O_5S$	285.0797	205.1225, 79.9575	Ibuprofen Sulfate
Organic acids ≤10 carbons				
M+H 138.0915	$C_8H_{11}NO$	138.0919	121.065, 93.0690, 91.0534	m-Tyramine
		Metabolites unique to neat urine		
Amino acids/ metabolites				
M-H147.0301	$C_5H_8O_5$	147.0293	129.0204,101.0261, 85.0322	Hydroxyglutamate
Nucleotides/ metabolites				
M-H 243.0631	$C_9H_{12}N_2O_6$	243.0630	200.0556	Uridine
M-H 195.0522	$C_7H_8N_4O_3$	195.0518	nd	Dimethyl uric acid

Table 3.2 continued				
M+H 183.0521	$C_6H_6N_4O_3$	183.0518	191.9606	Methyl uric acid
Vitamins				
M-H 135.0302	$C_4H_8O_5$	135.0293	nd	Threnoic acid
Dietary compounds				
M-H 221.0670	$C_8H_{14}O_7$	221.0661	175.0237, 157.0116	Ethyl glucuronide
M-H 189.0411	$C_8H_{14}OS_2$	189.0408	nd	Methialdol[1]
M+H 167.0573	$C_6H_6N_4O_2$	167.0569	110.0338	Methylxanthine
M-H 383.1201	$C_{14}H_{24}O_{12}$	383.1190	324.0712	Acetyl maltose
M-H 179.0571	$C_6H_{12}O_6$	179.0556	161.0463	Unidentified sugar
Organic acids ≤10 carbons				
M-H 161.0460	$C_6H_{10}O_5$	161.0450	nd	Unidentified organic acid
M-H 191.0195	$C_6H_8O_7$	191.0192	147.307,129.0195, 111.0088	Citric acid

Metabolite identities were confirmed with standards or Q-TOF fragmentation pattern. [1]Putative identity based upon accurate mass measurement of molecular ion and a knowledge of dietary and pharmaceutical intake in addition to the more common metabolites present in urine to make an informed decision as to the potential identity of metabolites where genuine standards or database fragmentation details are not available. nd = metabolite signal too weak for detection of fragments.

3.4.2 nUHPLC-nESI-TOF system stability

The variability of retention time for the nano platform was assessed by calculating the %CV for the top 100 most abundant metabolites (50 in positive nESI and 50 in negative nESI). All of these peaks returned a %CV for retention time variability of less than the 2%, a threshold suggested for metabolomics analysis (Theodoridis et al., 2012), (Figure 3.5). Those eluting early, typically before 7.5 mins showed the most variability due to the compounds being poorly retained and subsequently suffering from poor peak shape. The poor retention of these compounds is a result of the reversed phase stationary phase not being able to retain the metabolites on column. However, the peak shapes of early compounds peaks have been optimized by reconstituting samples in a high percentage of aqueous solvent, which we have observed to greatly improve peak shape compared to high organic solvent (data not shown). However, 96% of the peaks gave a %CV for retention time of less than 1%, and 73% of all peaks had a %CV of <0.2%, 10 times lower than the recommended threshold. The peaks between 10 and 15 mins with a %CV of between 0.3-0.6% were sulfated steroid metabolites which are observed to tail slightly on this column chemistry. In addition, the variation in peak area for the entire metabolome was assessed using the QC samples by calculating %CV for all peaks in the metabolome that are present in at least 80% of the samples (Want et al., 2010). This analysis revealed that the response of the nano platform was highly repeatable with 81% (+nESI) and 79% (-ESI) of peaks with a %CV of the mean peak area of <30%.

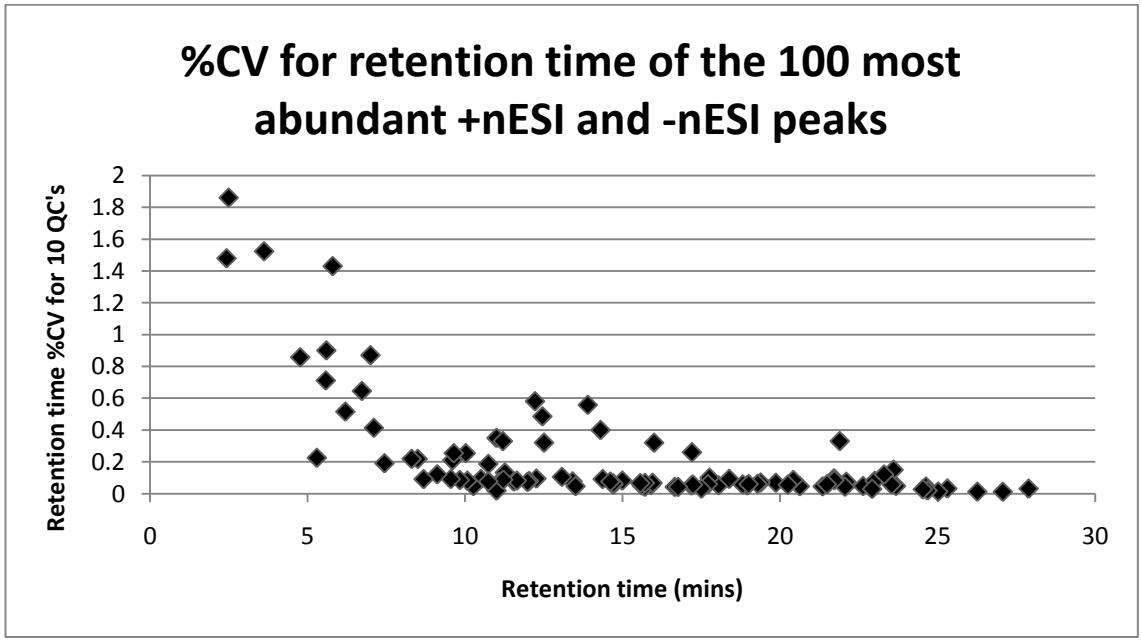


Figure 3.5: Retention time repeatability of nUHPLC-nESI-TOFMS analyses

Retention time variability calculated as %CV from the mean retention time of the 100 most abundant peaks (50 +nESI and 50 -nESI) present in the 10 QC injections. All values are < 1% for resolved peaks.

3.4.3 Comparison of UHPLC-ESI-TOFMS vs nUHPLC-nESI-TOFMS profiling.

The development of more sensitive high resolution TOF instruments using nanoflow-nanospray technologies show great promise for small molecule profiling (Chetwynd et al., 2014, Jones et al., 2014a). A decrease in column diameter and spray volume can result in significant improvements in sensitivity due to improved peak resolution, ionization efficiency and reduced ion suppression from matrix components (Liu et al., 2012, Gangl et al., 2001). In this study, direct comparison of datasets of urine profiles from conventional and nanospray platforms was not feasible using multivariate methods due to differences in retention time on the UHPLC versus nUHPLC columns. Instead, a comparison of the two platforms to detect metabolites discriminating between the male and female samples was performed as a surrogate marker for gains in sensitivity on the nano system.

SPE extracts were injected at 5 µL urine equivalence on each platform. OPLS-DA analysis of metabolomic profiles from the replicate SPE extracts of samples from male and female subjects was carried out for each MS platform, and metabolites unique for samples from either of the two individuals were detected by S-plot analysis. Concentrated SPE sample profiling by conventional UHPLC-ESI-TOFMS resulted in the detection of 23 metabolites which were exclusive to either the female or male subjects, namely bile acids and markers of diet or smoking (see Table 3.3). However, analysis by the nanospray platform resulted in detection of a further 19 metabolites unique to either the male or female extracts and these included compounds associated with steroid, eicosanoid, lipid signalling pathways, neurotransmitter and amino acid metabolism. As a result, the nUHPLC-nESI-TOFMS chromatograms were further interrogated for additional metabolites from these pathways, and which were not detected by the UHPLC-ESI-TOFMS platform (see Table 3.3). Unique to analysis by the nano platform revealed that products of the lipoxygenase pathway such as leukotriene E4 and hydroxyeicosatetraenoic acids (HETEs) were markers in (male) urine. These eicosanoids are mediators of inflammation and disease (Singh et al., 2010), however, other similar (prostaglandin) metabolites were not detected in the extracts of either male or female urine, possibly because markers of non-steroidal anti-inflammatory drugs (NSAIDs) were present in urine samples from both subjects indicating exposure to pharmaceuticals which inhibit cyclooxygenase activity and prostanoid biosynthesis. Progesterone, free and conjugated estrogens were identified by the nano platform as markers in female urine, and further interrogation of the chromatograms revealed a number of androgen and glucocorticoid metabolites also unique to analysis by the nano platform (see Figure 3.6 for examples and Table 3.3 for a full list).

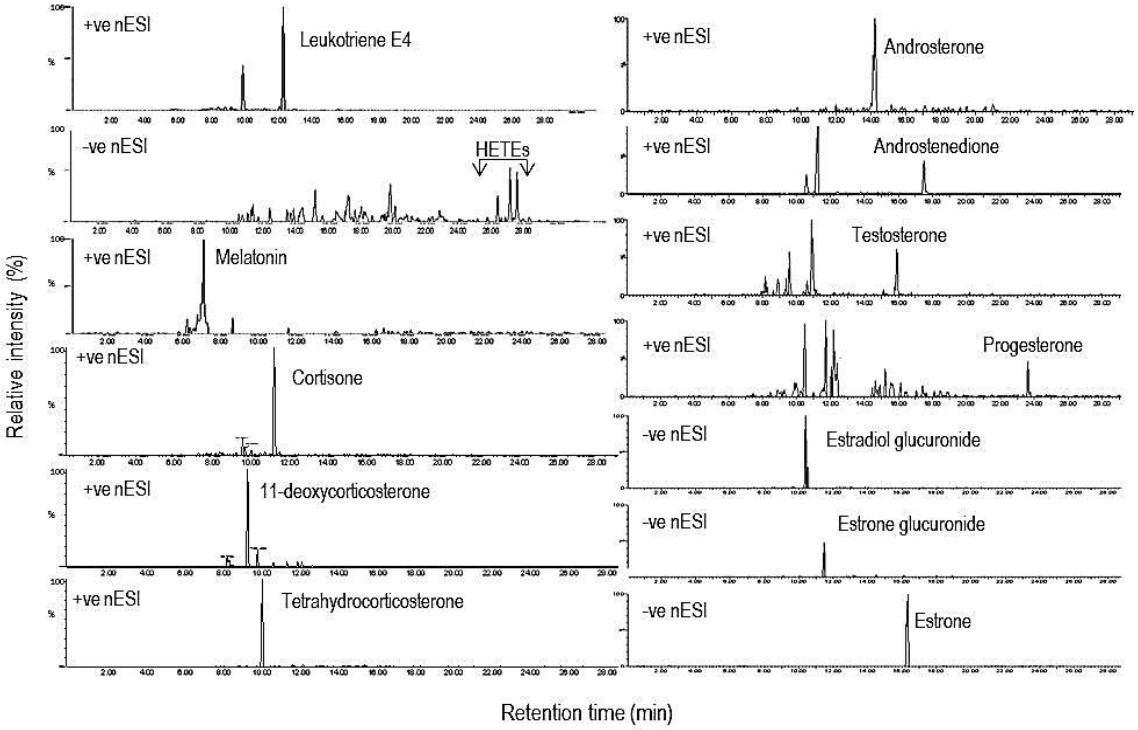


Figure 3.6: Examples of profiles of selected signalling metabolites detected in urine by nUHPLC-nESI-TOFMS analysis

Urine samples were profiled by nUHPLC-nESI-TOFMS and compounds identified by comparison to standards and fragmentation pattern. Selected ion chromatograms, at a mass resolution window of 2 ppm, are shown for estrogens, HETES, conjugated androgens an glucocorticoids in –ve ESI mode and progesterone, LTE4, melatonin, glucocorticoids and free androgens in +ve ESI mode. Metabolites corresponding to the additional signals in the ion chromatograms of progesterone, testosterone, androstenedione and LTE4 were not identified.

Many of these metabolites are typically present at very low levels in urine at concentrations between <0.5 nM and 3 nM/mM creatinine and not normally detected using current metabolomics methods (Bouatra et al., 2013). Free estrogens are also poorly ionized and not detected using the conventional ESI platform even after SPE concentration of urine samples. As sex steroids play a role in several cancers, fertility and cardiovascular disease and are also potential markers of endocrine disruption, the ability to detect these metabolites during metabolomics profiling constitutes an important step forward in understanding their role in disease processes (Gouveia et al., 2013). Other signalling metabolites detected by the nano platform included the serotonin metabolite methoxytryptamine (Table 3.3), melatonin which regulates circadian activity (Blask et al., 2011), and the neurotransmitter dopamine. The detection of these additional metabolites that were unique to the nano platform highlights the gains in sensitivity and this could be due to a number of factors. In nESI, the increased rate of desolvation and ionization efficiency afforded by the fine spray from the nanospray emitter enables a greater proportion of molecules entering the source to become ionized and thus allowing the detection of low abundance compounds in the sample extract. In addition,

compared with conventional UHPLC, stability of retention time on the nanoflow platform was achieved with a 20x lower concentration of the acid buffer (FA), and thus this would allow more sensitive analysis of acidic or phenolic compounds detected in the -nESI mode. However, a maximum injection volume of 0.5 µL was required to ensure acceptable peak shape in the nanoflow system and hence pre-concentration of samples by SPE is vital to allow a large urine equivalence to be injected on the nanoflow column in order to increase detection levels of many metabolites In addition, the more extensive sample cleanup is necessary with the nanoflow system in order to remove salts and proteins as even very fine particulates can cause blockages of the column or nESI emitter causing poor nanospray stability and fluctuations in instrument sensitivity.

Table 3.3: Comparison of UHPLC-ESI-TOFMS and nUHPLC-nESI-TOFMS platforms for detection of discriminatory metabolites

Observed m/z	Structure	Theoretical mass	Fragments	Unique to male /female samples	Identity
Discriminatory metabolites of subject detected by UHPLC-ESI-TOFMS					
Neurotransmitter/ metabolites					
M-H 359.0979	$C_{15}H_{20}O_{10}$	359.0978	175.0230, 113.0224	Female	Methoxy-hydroxyphenylglycol glucuronide[1]
M-H 263.0231	$C_9H_{12}O_7S$	263.0225	183.0682	Male	Methoxy hydrophenylglycol sulfate
Bile acids and metabolites					
M-H 391.2843	$C_{24}H_{40}O_4$	391.2848	nd	Female	Chenodeoxycholic acid
M-H 498.2886	$C_{26}H_{45}NO_6S$	498.2889	nd	Male	Taurodeoxycholic acid
Lipid metabolites					
M+H 542.2975	$C_{24}H_{47}NO_{10}S$	542.2990	463.092	Female	Psychosine sulfate
Dietary compounds					
M-H 445.1129	$C_{22}H_{22}O_{10}$	445.1135	nd	Female	Unidentified glucoside[1]
M+H 304.0930	$C_{16}H_{15}O_6$	304.0947	nd	Female	Unidentified flavanoid[1]
M+H 401.1791	$C_{19}H_{28}O_9$	401.1812	nd	Female	Corchoionoside B[1]
M-H 557.0580	$C_{22}H_{22}O_{15}S$	557.0601	nd	Female	Unidentified sulfated glucoside[1]
M-H 387.1831	$C_{22}H_{28}O_6$	387.1808	nd	Female	Unidentified dietary sugar[1]
M-H 381.0291	$C_{16}H_{14}O_9S$	381.0280	nd	Female	Hesperetin sulfate[1]
M-H 477.1040	$C_{22}H_{22}O_{12}$	477.1033	nd	Female	Unidentified glucoside[1]
M-H 321.1554	$C_{14}H_{26}O_8$	321.1549	nd	Female	Butyl hydroxyl butyrate glucoside[1]

Table 3.3 continued					
M-H 395.0069	$C_{12}H_{19}Cl_3O_8$	395.0067	359.0368	Male	Sucralose
M-H 217.0185	$C_8H_{10}O_5S$	217.0171	nd	Female	Tyrosol sulfate[1]
M-H 571.1451	$C_{28}H_{28}O_{13}$	571.1452	nd	Male	Triacetylglycitin[1]
M-H 313.0552	$C_{13}H_{14}O_9$	313.0560	175.0216, 113.0267	Male	Salicylate glucuronide[1]
M-H 427.1959	$C_{21}H_{32}O_9$	427.1968	nd	Male	Taraxacolide glucopyranoside[1]
Smoking related compounds					
M+H 177.1029	$C_{10}H_{12}N_2O$	177.1028	80.049	Female	Cotinine[1]
M+H 193.0974	$C_{10}H_{12}N_2O_2$	193.0977	143.117	Female	Cotinine N-Oxide
M+H 163.1238	$C_{10}H_{14}N_2$	163.1235	132.0785, 130.0785	Female	Nicotine
Organic acids ≤10 carbons					
M-H 211.0607	$C_{10}H_{12}O_5$	211.0606	123.046	Female	Unidentified acid[1]
Organic acids >10 carbons					
M-H 224.0599	$C_{11}H_{13}O_5$	224.0685	nd	Male	Unidentified acid[1]
Additional discriminatory metabolites of subject detected by nUHPLC-nESI-TOFMS:					
Amino acids/ metabolites					
M-H165.0547	$C_9H_{10}O_3$	165.0552	147.042, 119.0504	Female	Phenyllactic acid
Neurotransmitter/ metabolites					
M+H 191.1185	$C_{11}H_{14}N_2O$	191.1184	174.092, 159.068	Male	5-Methoxytryptamine
Conjugated androgens					
M-H 383.1525	$C_{19}H_{28}O_6S$	383.1528	96.9595, 79.9571	Female	Unidentified sulfated androgen
M-H 467.2641	$C_{25}H_{40}O_8$	467.2645	175.0240, 157.0141, 113.0241	Male	Unidentified androgen glucuronide[1] (retention time 12.89 mins)

Table 3.3 continued					
Free estrogens					
M-H 269.1545	$C_{18}H_{22}O_2$	269.1542	221.1563, 145.0656	Female	Estrone
Conjugated estrogens					
M-H 445.1871	$C_{24}H_{30}O_8$	445.1862	175.0231, 157.1220, 113.0231	Female	Estrone glucuronide
M-H 477.2133	$C_{25}H_{34}O_9$	477.2125	175.0223, 113.0225	Female	Methoxy-estradiol-glucuronide[1]
Progestogens					
M+H 315.2332	$C_{21}H_{30}O_2$	315.2324	nd	Female	Progesterone
M-H 395.1888	$C_{21}H_{32}O_5S$	395.1882	79.9579	Female	Pregnenolone sulfate[1]
Eicosanoids					
M-H 319.2275	$C_{20}H_{32}O_3$	319.2273	285.2281, 303.2351	Female	Unidentified HETE
M+H 440.2477	$C_{23}H_{37}NO_5S$	440.2471	319.2250, 189.165	Male	Leukotriene E4
Lipid metabolites					
M+H 282.2790	$C_{18}H_{35}NO$	282.2797	nd	Male	Oleamide[1]
Dietary compound					
M-H 533.1382	$C_{25}H_{26}O_{13}$	533.1373	nd	Female	Unidentified glucoside
M-H 367.1421	$C_{18}H_{24}O_8$	367.1393	nd	Female	Unidentified glucoside
M-H 204.0665	$C_{11}H_{11}NO_3$	204.0661	160.0763,132.0811,130.0652, 117.07, 103.0546	Female	Cinnamoylglycine
M+H 261.1231	$C_{14}H_{16}N_2O_3$	261.1239	nd	Male	Maculosin L,L-Cyclo(leucylprolyl)[1]
M-H 359.1141	$C_{19}H_{20}O_7$	359.1131	nd	Male	Unidentified flavonoid[1]
Organic acids ≤10 carbons					
M-H 147.0440	$C_9H_8O_2$	147.0446	103.0554	Male	Cinnamic acid

Table 3.3 continued					
Organic acids >10 carbons					
M-H 427.1840	$C_{18}H_{28}N_4O_8$	427.1829	nd	Male	Pyridinoline[1]
Additional metabolites detected in both subjects and unique to nUHPLC-nESI-TOFMS:					
Amino acids/ metabolites					
M+H 205.0973	$C_{11}H_{12}N_2O_2$	205.0977	188.0625	-	Tryptophan
M+H 225.0884	$C_{10}H_{12}N_2O_4$	225.0875	208.0620	-	Hydroxykynurenine
M-H 159.0293	$C_6H_8O_5$	159.0293	59.014	-	Oxoadipic acid
M+H 175.0721	$C_6H_{10}N_2O_4$	175.0719	nd	-	Formiminoglutamic acid[1]
M+H 145.1332	$C_7H_{16}N_2O$	145.1341	nd	-	Acetylcadaverine[1]
M-H 143.0347	$C_6H_8O_4$	143.0344	nd	-	Methylglutaconic acid[1]
M-H 138.0194	$C_6H_5NO_3$	138.0191	94.0320	-	Hydroxypicolinic acid
Neurotransmitter/ metabolite					
M+H 154.0841	$C_8H_{11}NO_2$	154.0868	137.0495, 113.0630	-	Dopamine
M+H 233.1295	$C_{13}H_{16}N_2O_2$	233.1290	nd	-	Melatonin
Free androgens					
M+H 289.2162	$C_{19}H_{28}O_2$	289.2168	nd	-	Testosterone
M+H 287.2016	$C_{19}H_{26}O_2$	287.2011	nd	-	Androstenedione
Conjugated androgens					
M-H 467.2641	$C_{25}H_{40}O_8$	467.2645	175.0240, 157.0141, 113.0241	-	Unidentified androgen glucuronide[1] (retention time 15.53 mins)
Conjugated estrogens					
M-H 447.2021	$C_{24}H_{32}O_8$	447.2019	175.0237, 157.1229, 113.0232	-	Estradiol glucuronide

Table 3.3 continued					
Glucocorticoids					
M+H 361.2021	$C_{21}H_{28}O_5$	361.2015	326.16, 301.165	-	Cortisone
M+H 331.2278	$C_{21}H_{30}O_3$	331.2273	nd	-	11-deoxycorticosterone
M+H 335.2581	$C_{21}H_{34}O_3$	335.2586	nd	-	Tetrahydrodeoxycorticosterone
Bile acid/ metabolites					
M-H 377.2699	$C_{23}H_{38}O_4$	377.2848	nd	-	Apocholic acid[1]
M-H 583.3128	$C_{30}H_{48}O_{11}$	583.3118	407.2796, 175.0234, 157.1230, 113.0233	-	Cholic acid glucuronide
Lipid metabolites					
M+H 300.2902	$C_{16}H_{37}NO_2$	300.2903	282.2767	-	Sphingosine
M-H 295.2275	$C_{18}H_{22}O_3$	295.2273	277.2170, 195.1440, 113.0960, 71.0150	-	Hydroxyoctadecadienoic acid
M-H 199.0972	$C_{10}H_{16}O_4$	199.0970	nd	-	Decanoic acid[1]
M-H 171.1386	$C_{10}H_{20}O_2$	171.1385	68.9971	-	Capric acid
M-H 155.1075	$C_9H_{16}O_2$	155.1072	nd	-	Hydroxynonenal[1]
M+H 170.0450	$C_7H_7NO_4$	170.0453	nd	-	Furoylglycine[1]
M+H 232.1181	$C_{10}H_{17}NO_5$	232.1185	nd	-	Suberylglycine
M+H 158.0813	$C_7H_{11}NO_3$	158.0817	nd	-	Methylcrotonylglycine[1]
M-H 173.0813	$C_8H_{14}O_4$	173.0814	111.0810	-	Suberic acid
Dietary compounds					
M-H 165.0551	$C_9H_{10}O_3$	165.0552	121.0655	-	Methoxyphenylacetic acid
M-H 307.0283	$C_{14}H_{12}O_6S$	307.0276	nd	-	Resveratrol sulfate [1]
M-H 391.1230	$C_{20}H_{22}O_8$	391.1393	nd	-	Resveratrol glucoside[1]
M-H 403.1028	$C_{20}H_{20}O_9$	403.1029	175.0229, 113.0225	-	Resveratrol glucuronide[1]

Table 3.3 continued					
M-H 155.1437	$C_{10}H_{20}O$	155.1436	nd	-	Menthol[1]
M-H 273.0765	$C_{15}H_{14}O_5$	273.0763	167.0340, 119.0490	-	Phloretin
M-H 253.0505	$C_{15}H_{10}O_4$	253.0501	91.0180	-	Daidzein
M-H 379.1392	$C_{19}H_{24}O_8$	379.1393	nd	-	Unidentified glucoside[1]
M+H 295.1300	$C_{14}H_{18}N_2O_5$	295.1294	nd	-	Aspartame[1]
M-H 297.1130	$C_{18}H_{18}O_4$	297.1127	nd	-	Enterolactone[1]
M-H 255.0660	$C_{15}H_{12}O_4$	255.0657	nd	-	Dihydrodaidzein[1]
M-H 283.0598	$C_{16}H_{12}O_5$	283.0606	268.0380	-	Glycitein
M-H 271.0612	$C_{15}H_{12}O_5$	271.0606	nd	-	Dihydrogenistein[1]
M+H 179.1065	$C_{11}H_{16}O_2$	179.1072	nd	-	Butylated hydroxyanisole[1]
M-H 137.0609	$C_8H_{10}O_2$	137.0603	119.0501	-	Tyrosol
M-H 301.1550	$C_{18}H_{22}O_4$	301.1440	nd	-	Enterodiol[1]
M-H 115.0760	$C_6H_{12}O_2$	115.0759	nd	-	Caproic acid
M-H 227.1291	$C_{12}H_{20}O_4$	227.1283	183.1380, 165.1280	-	Traumatic acid
Free pharmaceuticals					
M-H 205.1233	$C_{13}H_{18}O_2$	205.1229	161.1330, 154.9740	-	Ibuprofen
M+H 152.1068	$C_9H_{13}NO$	152.1075	134.0961, 117.0695	-	Phenylpropanolamine
Organic acids ≤10 carbons					
M+H 188.1760	$C_9H_{21}N_3O$	188.1763	171.1490, 117.1020, 100.0750	-	Acetylspermidine

Compound identities were confirmed by comparison of fragments with databases or standards. [1]Putative identity based upon accurate mass measurement of molecular ion and a knowledge of dietary and pharmaceutical intake in addition to the more common metabolites present in urine to make an informed decision as to the potential identity of metabolites where genuine standards or database fragmentation details are not available. nd = metabolite signal too weak for detection of fragments

3.4.4 Comparison of weak anion exchange vs strong cation exchange SPE using nUHPLC-nESI-TOFMS profiling.

In the above studies, the SPE methodology used to extract urine samples was a combination of two SPE cartridge phases; urine samples were first extracted using X-AW SPE which is a mixed mode polymeric and weak anion exchange phase with an affinity for hydrophilic, hydrophobic neutral species as well as the strongly anionic compounds present in urine. The urine eluate was from the X-AW SPE was then passed through a mixed mode strong cation exchange phase (X-C) to extract weak cationic species which would have been poorly retained by the X-AW phase. Compounds were separately eluted from the two SPE phases and aliquots of both the extracts combined for MS analyses. In further work, the contribution of the X-C SPE phase to the metabolite profiles extracted by X-AW was evaluated. Initial studies indicated that after X-AW extraction, an additional extraction by the X-C phase resulted in detection of a significant number of markers (features) after nUHPLC-nESI-TOFMS analysis (Figure 3.7). The markers extracted by X-AW and the X-C cartridges were qualitatively compared using OPLS-DA models and an S-plot analysis of the loading variables. Many of the MS signals extracted by X-C were also present at higher concentrations in X-AW extracts. However, S-plot analysis revealed that 57 metabolites were unique to the X-C phase and not detected at all in X-AW extracts (Table 3.4). These included carnitines which play important roles in lipid metabolism, nucleotides, amino acids and nicotine metabolites. Many of these metabolites are cationic species containing nitrogen groups which would have been poorly retained on the X-AW chemistry due to electrostatic repulsion between the positively charged analyte and the X-AW phase. nUHPLC-nESI-TOFMS analysis of a combined extract from the X-AW and X-C phases revealed a lower number of markers compared with X-AW extracts alone, particularly in –nESI mode (Figure 3.7). Extracts from individual SPE phases were analysed separately by nUHPLC-nESI-TOFMS, and these datasets were also compared with an analysis of the combined aliquots of extracts from both SPE phases to determine whether information was lost by analysing a combined SPE extract. OPLS-DA modelling of the datasets of extracts from combined and individual SPE revealed that 55 metabolites were not detected in the combined extracts compared to analysis of the individual SPE phases (Table 3.5). These metabolites included amino acids, vitamins, lipids, neurotransmitters, pharmaceuticals and short chain organic acids. A likely explanation for their lack of detection in the combined SPE extract was due to ion suppression resulting from co-eluting compounds present in either the X-AW or X-C extracts. This result indicates that to extract maximum metabolite information

from the urine samples, the extracts from the anionic and cationic SPE phases should be analysed separately.

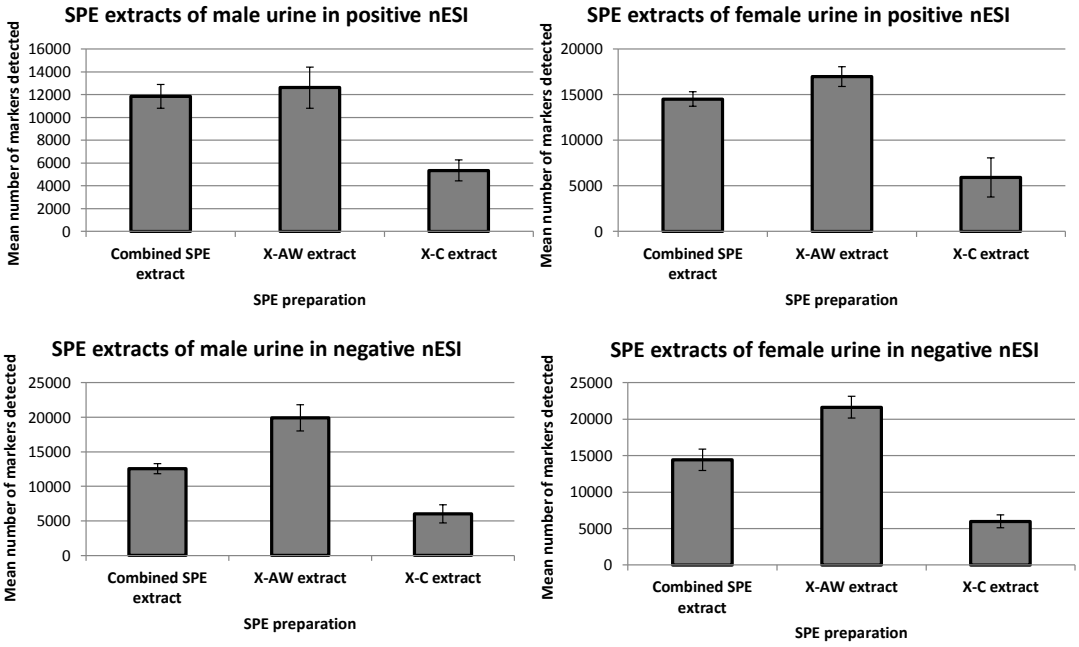


Figure 3.7: Number of markers detected in profiles of individual SPE preparations compared with profiles from combined SPE samples

Number of markers in extracts from X-AW and/or X-C SPE phases are given as mean ±SD (n=5) after removal of workup blank signals. Graphs A and B markers detected in positive nESI modes and Graphs C and D markers detected in negative nESI mode for the male and female samples.

Table 3.4: Metabolites markers unique to the strong cation exchange SPE cartridge

Experimental mass	Structure	Theoretical mass	Q-TOF Fragments	Identity
Amino acids/ metabolites				
M+H 156.0769	$C_6H_9N_3O_2$	156.0773	110.0719	Histidine
M+H 195.0769	$C_9H_{10}N_2O_3$	195.0770	149.071, 120.0450	Aminohippuric acid
M-H 192.0664	$C_{10}H_{11}NO_3$	192.0661	148.075, 91.057	Methylhippuric acid
M+H 170.9250	$C_7H_{11}N_3O_2$	170.0930	126.1021, 109.0751, 96.0683	Methyl histidine
M+H 166.0861	$C_9H_{11}NO_2$	166.0868	120.0801	Phenyl alanine
M-H 173.0918	$C_7H_{12}N_2O_3$	173.0926	116.0710, 70.0660	Glycylproline
M+H 182.0824	$C_9H_{11}NO_3$	182.0817	165.0560 136.0764	Tyrosine
M+H 291.1314	$C_{10}H_{18}N_4O_6$	291.1305	nd	Argininosuccinic acid[1]
M+H 203.1500	$C_8H_{18}N_4O_2$	203.1508	nd	Symmetric dimethylarginine[1]
M+H 295.1292	$C_{14}H_{18}N_2O_5$	295.1294	nd	Glutamylphenylalanine[1]
M-H 269.0603	$C_{11}H_{14}N_2O_4S$	269.0596	nd	Hydroxyphenylacetothiohydroximoyl cysteine[1]
M+H 139.0502	$C_6H_6N_2O_2$	139.0508	121.0400	Urocanic acid
M+H 118.0860	$C_5H_{11}NO_2$	118.0868	101.0600	Aminopentanoic acid
M+H 145.1332	$C_7H_{16}N_2O$	145.1341	nd	N-Acetylcadaverine[1]
M+H 138.0919	$C_8H_{11}NO$	138.0919	121.065, 93.0690, 91.0534	m-Tyramine
Carnitines				
M+H 162.1121	$C_7H_{15}NO_3$	162.1130	nd	Carnitine
M+H 232.1549	$C_{11}H_{21}NO_4$	232.1549	173.0808	Butyrlcarnitine
M+H 232.1549	$C_{11}H_{21}NO_4$	232.1549	173.0808	Isobutyrlcarnitine
M+H 244.1547	$C_{12}H_{21}NO_4$	244.1549	nd	Ethylacrylcarnitine[1]

Table 3.4 continued				
M+H 246.1706	$C_{12}H_{23}NO_4$	246.1705	187.0985	Methylbutyroylcarnitine
Neurotransmitter/ metabolites				
M+H 191.1185	$C_{11}H_{14}N_2O$	191.1184	174.092, 159.068	Methoxytryptamine
M+H 178.0858	$C_{10}H_{11}NO_2$	178.0868	nd	Hydroxytryptophol[1]
M+H 249.1231	$C_{13}H_{16}N_2O_3$	249.1239	190.0858	Hydroxymelatonin
M+H 285.1924	$C_{13}H_{24}N_4O_3$	285.1927	nd	Melanostatin[1]
M+H 152.1068	$C_9H_{13}NO$	152.1075	134.0961, 117.0695	Phenylpropanolamine
M+H 176.0715	$C_{10}H_9NO_2$	176.0712	nd	Hydroxyindoleacetaldehyde[1]
Nucleotides/metabolites				
M+H 166.0720	$C_6H_7N_5O$	166.0729	149.0459	Methylguanine
M-H 310.1163	$C_{12}H_{17}N_5O_5$	310.1151	nd	Dimethylguanosine[1]
M+H 268.1066	$C_{10}H_{13}N_5O_4$	268.1046	136.0615	Deoxyguanosine
M+H 282.1199	$C_{11}H_{15}N_5O_4$	282.1202	136.0620, 119.0350	Methyladenosine
M-H 167.0223	$C_5H_4N_4O_3$	167.0205	124.0163	Uric acid
Vitamins				
M+H 245.0953	$C_{10}H_{16}N_2O_3S$	245.0960	227.0860	Biotin
M-H 213.0699	$C_9H_{14}N_2O_2S$	213.0698	nd	Methyl bisnorbiotinyl ketone[1]
M-H 215.0483	$C_8H_{12}N_2O_3S$	215.0490	nd	Bisnorbiotin[1]
M+H 220.1185	$C_9H_{17}NO_5$	220.1186	202.1088	Pantothenic acid
Lipid metabolites				
M+H 185.0815	$C_5H_{15}NO_4P$	185.0817	nd	Phosphorylcholine[1]
Dietary compounds				

Table 3.4 continued				
M-H 195.0518	$C_6H_{12}O_7$	195.0505	129.0197, 75.0090	Gluconic acid
M-H 165.0428	$C_6H_6N_4O_2$	165.0413	nd	Methylxanthine[1]
M+H 153.0665	$C_7H_8N_2O_2$	153.0664	nd	Methyl pyridonecarboxamide[1]
Smoking related compounds				
M+H 177.1029	$C_{10}H_{12}N_2O$	177.1028	80.049	Cotinine
M+H 193.0974	$C_{10}H_{12}N_2O_2$	193.0977	143.117	Cotinine N-Oxide
M+H 149.1074	$C_9H_{12}N_2$	149.1079	nd	Nornicotine[1]
M+H 193.0976	$C_{10}H_{12}N_2O_2$	193.0977	nd	Hydroxycotinine[1]
M+H 179.1176	$C_{10}H_{14}N_2O$	179.1184	nd	Nicotine-N-oxide[1]
M+H 163.1227	$C_{10}H_{14}N_2$	163.1235	146.095	Anabasine
Organic acids ≤10 carbons				
M+H 130.0862	$C_6H_{11}NO_2$	130.0868	84.0814, 56.0490	Pipecolic acid
M-H 138.0194	$C_6H_5NO_3$	138.0191	94.032	Hydroxypicolinic acid
M-H 115.0028	$C_4H_4O_4$	115.0031	71.0140	Maleic acid
M-H 182.0480	$C_5H_{13}NO_4S$	182.0487	nd	Choline sulfate[1]
M-H 111.0085	$C_5H_4O_3$	111.0082	67.0190	Furoic acid
M+H 175.0721	$C_6H_{10}N_2O_4$	175.0719	nd	Formiminoglutamic acid[1]
M+H 149.0929	$C_5H_{11}N_3O_2$	146.0930	104.0700, 87.0450, 86.0600	Guanidinobutanoic acid
M+H 106.0868	$C_4H_{11}NO_2$	106.0868	88.0760	Diethanolamine
M+H 144.0668	$C_4H_7N_3O$	144.0667	nd	Creatinine

Table 3.4 continued				
Other				
M-H 308.0782	$C_{14}H_{15}NO_7$	308.0770	175.0243, 157.0147, 113.0241	Inodxyl glucuronide[1]

Metabolite identities were confirmed with standards or Q-TOF fragmentation pattern. [1]Putative identity based upon accurate mass measurement of the molecular ion and a knowledge of dietary and pharmaceutical intake in addition to the more common metabolites present in urine to make an informed decision as to the potential identity of metabolites where genuine standards or database fragmentation details are not available. nd = metabolite signal too weak for detection of fragments.

Table 3.5: Metabolites not detected from analysis of combined SPE extracts compared with extracts from X-AW and X-C phases analysed separately

Experimental mass	Structure	Theoretical mass	Fragments	Presence in SPE phase	Identity
Amino acid/ metabolites					
M-H 193.0608	$C_9H_{10}N_2O_3$	193.0613	149.071, 120.045	X-C	Aminohippuric acid
M-H 192.0664	$C_{10}H_{11}NO_3$	192.0661	148.075, 91.057	X-C	Methylhippuric acid
M-H 129.0546	$C_6H_{10}O_3$	129.0552	nd	X-AW	Ketoleucine[1]
M+H 295.1292	$C_{14}H_{18}N_2O_5$	295.1294	nd	X-C	Glutamylphenylalanine[1]
M+H 291.1314	$C_{10}H_{18}N_4O_6$	291.1305	nd	X-C	Argininosuccinic acid[1]
M+H 138.0915	$C_8H_{11}NO$	138.0919	121.065, 93.0690, 91.0534	X-C	m-Tyramine
M+H 304.1296	$C_{15}H_{17}N_3O_4$	304.1297	nd	Both	Indoleacetyl glutamine[1]
M+H 203.1502	$C_8H_{18}N_4O_2$	203.1508	nd	X-C	Symmetric dimethylarginine[1]
M+H 118.0861	$C_5H_{11}NO_2$	118.0868	101.06	X-C	Aminopentanoic acid
M+H 241.1296	$C_{10}H_{16}N_4O_3$	241.1301	nd	Both	Homocarnosine[1]
M+H 130.0862	$C_6H_{11}NO_2$	130.0868	84.0814, 56.049	X-C	Pipecolic acid
M+H 146.0823	$C_6H_{11}NO_3$	146.0817	nd	X-AW	Acetamidobutanoic acid[1]
M+H 118.0652	C_8H_7N	118.0657	91.0554	Both	Indole
M-H 308.0782	$C_{14}H_{15}NO_7$	308.0770	175.0243, 157.0147, 113.0241	X-C	Indoxyl glucuronide[1]
Carnitines					
M+H 218.1399	$C_{10}H_{19}NO_4$	218.1392	159.0628	X-C	Propionylcarnitine
Neurotransmitter/ metabolites					
M+H 198.1131	$C_{10}H_{15}NO_3$	198.1130	nd	X-AW	Metanephrine[1]
M+H 285.1924	$C_{13}H_{24}N_4O_3$	285.1927	nd	X-C	Melanostatin[1]
M+H 176.0712	$C_{10}H_9NO_2$	176.0712	nd	X-C	Hydroxyindoleacetaldehyde[1]
M+H 141.0654	$C_6H_8N_2O_2$	141.0664	nd	Both	Methylimidazoleacetic acid[1]
Vitamins					
M+H168.0665	$C_8H_9NO_3$	168.0661	150.0548, 127.0119	X-AW	Pyridoxal (Vit B6)
M+H 184.0610	$C_8H_9NO_4$	184.0606	166.0507, 148.0397	Both	Pyridoxic acid (Vit B6)
M+H 245.0953	$C_{10}H_{16}N_2O_3S$	245.0960	227.086	X-AW	Biotin
M-H 215.0483	$C_8H_{12}N_2O_3S$	215.0490	nd	X-C	Bisnorbiotin[1]
Lipid metabolites					
M+H 300.2902	$C_{18}H_{37}NO_2$	300.2903	282.2767	X-AW	Sphingosine
M+H 256.2644	$C_{16}H_{33}NO$	256.2640	nd	Both	Palmitic amide[1]
M+H 170.0450	$C_7H_7NO_4$	170.0453	nd	X-C	Furoylglycine[1]

Table 3.5 continued					
M+H 158.0813	$C_7H_{11}NO_3$	158.0817	nd	X-AW	Methylcrotonylglycine[1]
M+H 232.1181	$C_{10}H_{17}NO_5$	232.1185	nd	X-AW	Suberylglycine[1]
M-H 227.2013	$C_{14}H_{28}O_2$	227.2011	nd	Both	Myristic acid[1]
M+H 181.0606	$C_8H_8N_2O_3$	181.0613	135.0556	X-AW	Nicotinuric acid
Dietary compounds					
M+H 138.0555	$C_7H_7NO_2$	138.0549	94.0661	Both	Trigonelline
M-H 221.0665	$C_8H_{14}O_7$	223.0818	175.0239, 157.0137, 113.0239	X-AW	Ethyl glucuronide[1]
M-H 149.0095	$C_4H_6O_6$	149.0086	103.0034, 87.009, 72.993, 59.015	Both	Tartaric acid
M-H 133.0138	$C_4H_6O_5$	133.0137	115.004, 71.0145	Both	Malic acid
Smoking related compounds					
M+H 163.1227	$C_{10}H_{14}N_2$	163.1235	146.095	X-C	Anabasine
Free pharmaceuticals					
M+H 152.0701	$C_8H_9NO_2$	152.0712	80.049, 110.061	Both	Acetaminophen
M-H 152.0350	$C_7H_7NO_3$	152.0348	nd	X-AW	Aminosalicylic acid[1]
	Conjugated pharmaceuticals				
M+H 328.1043	$C_{14}H_{17}NO_8$	328.1032	175.0243,113.0239, 152.0701	X-AW	Acetaminophen Glucuronide[1]
Organic acid ≤10 carbons					
M-H 115.0028	$C_4H_4O_4$	115.0031	71.014	X-C	Maleic acid
M+H 168.0302	$C_7H_5NO_4$	168.0297	nd	X-AW	Quinolinic acid[1]
M+H 449.2554	$C_{18}H_{36}N_6O_5S$	449.2546	nd	Both	Glutathionylaminopropylcadaverine[1]
M-H 145.0136	$C_5H_6O_5$	145.0137	101.0241, 57.0354	Both	Oxoglutaric acid
M-H 147.0303	$C_5H_8O_5$	147.0293	129.0191, 103.0406, 85.0296, 57.0361	Both	Hydroxyglutaric acid
M-H 111.0085	$C_5H_4O_3$	111.0082	67.019	X-C	Furoic acid
M-H 129.0191	$C_5H_6O_4$	129.0188	85.03	Both	Glutaconic acid
M+H 146.1658	$C_7H_{19}N_3$	146.1657	nd	X-AW	Spermidine[1]
M-H 151.0397	$C_8H_8O_3$	151.0395	107.5	Both	Cresotinic acid
M+H 114.0659	$C_4H_7N_3O$	114.0667	nd	Both	Creatinine
M+H 106.0868	$C_4H_{11}NO_2$	106.0868	88.076	X-C	Diethanolamine

Metabolite identities confirmed with standards or comparison of QTOF fragmentation patterns, [1]Putative identity based upon accurate mass measurement of the molecular ion and a knowledge of dietary and pharmaceutical intake in addition to the more common metabolites present in urine to make an informed decision as to the potential identity of metabolites where genuine standards or database fragmentation details are not available. nd= metabolite signal too weak for detection of fragments

A summary of the additional metabolites detected in this study across a wide range of chemical classes for the different sample preparation techniques and the two analytical platforms is presented in Table 3.6. It illustrates that by diluting urine samples less metabolites are detected, and that the SPE methodology allows the detection of more metabolites than the other sample preparation methods. However, analysis by nanoflow-nanospray platform using newly developed methodologies for metabolomics, allowed detection of many more metabolites and importantly many of which were in additional compound classes, including sex steroids, glucocorticoids, eicosanoids and neurotransmitters. Metabolomic analyses require high throughput of samples in order to obtain sample sizes needed for robust biomarker discovery. Often high throughput approaches result in a compromise in analytical sensitivity and subsequent metabolome coverage. The methods developed in this study require increased effort in terms of sample processing and analysis but result in significant improvements in the detection of very low abundance metabolites in addition to those traditionally difficult to ionize. This enables more metabolic pathways to be monitored using a single analytical approach which could potentially lead to more in-depth analysis of the role of metabolites in disease states, and the identification of biomarkers using non-invasive sampling.

Table 3.6: Summary of the compound classes detected using the different sample preparation methodologies and analytical platforms

	Conventional UHPLC-ESI-TOFMS			nUHPLC-nESI-TOFMS	nUHPLC-nESI-TOFMS
Compound class	**Diluted**	**Neat[1]**	**Combined SPE[2]**	**Combined SPE[3]**	**Individual SPE[4]**
Amino acids/ metabolites	+	+3	+2	+8	+14
Carnitines	+	+	+1	+	+1
Neurotransmitter/ metabolites	nd	+1	+1	+3	+4
Nucleotides/ metabolites	+	+	-3	+2	+4
Vitamins	+	+2	+	+	+4
Free androgens	nd	nd	nd	+2	+
Conjugated androgens	+	+2	+	+3	+
Free estrogens	nd	nd	nd	+1	+
Conjugates estrogens	nd	nd	nd	+3	+
Progestogens	nd	nd	nd	+2	+
Glucocorticoids	+	+1	+	+3	+
Eicosanoids	nd	nd	nd	+2	+
Bile acids and metabolites	+	+3	+6	+2	+
Lipid metabolites	+	+	+3	+10	+7
Dietary compounds	+	+3	-2	+23	+4
Smoking related compounds	+	+2	+2	+	+1
Free pharmaceuticals	+	+	+2	+2	+2
Conjugated pharmaceuticals	+	+	+1	+	+1
Organic acid ≤10 carbons	+	+5	-1	+2	+11
Organic acid >10 carbons	nd	+1	+	+1	+

The numbers of metabolites shown within each class were detected relative to the preceding sample preparation and analytical technique. + = detected/detected at same level as previous preparation/analytical technique; nd = none detected; *+number=* net increase in metabolites detected; *-number=* net decrease in metabolites detected. Figures compiled from data presented in: [1]Table 3.1; [2]Table 3.2; [3]Table 3.3; [4]Table 3.5.

3.5 Conclusion

This is the first study to introduce solid phase extraction sample preparation and direct nUHPLC-nESI-TOFMS to the field of untargeted urine metabolomic analysis and thus enable the profiling of low abundance components of the metabolome including signalling molecules important in understanding disease processes. The identification of low abundant metabolites in complex extracts of urine as part of a general non-targeted profiling methodology provides a step change in metabolomics analysis and a significant enhancement of coverage of the urinary metabolome

Chapter 4: A study on the effect of different normalisation methods for the correction of variability in urine concentration

4.1 Abstract

The normalisation of urine metabolomics data is of particular importance due to the large variation in urine concentration between voids. This differential dilution of urine may prevent, or lead to, incorrect associations between metabolite profiles and disease cases. To date, many studies normalise metabolomics data sets following LC-MS analysis and this limits normalisation to metabolites with a concentration that falls within the dynamic range of the mass spectrometer. The metabolites lost through dilution in the more dilute samples or saturate the detector in more concentrated samples are poorly normalised if at all. This effect was noted in the current study when post-analysis normalisation to creatinine, osmolality and mass spectrum total signal (MSTS) were investigated. To counter this, a pre-analysis normalisation method using osmolality was developed whereby samples were dried down following SPE extraction and reconstituted in a volume to give equal osmolality for all samples prior to LC-MS analysis. This resulted in improved scores plot clustering and increased the total number of common peaks and mean peak area repeatability of these peaks between dilution factors within individuals. Further improvements were observed when the pre-analysis normalisation to osmolality was combined with a post-analysis mass spectrum total useful signal (MSTUS) normalisation. Future adoption of such normalisation methods may aid the discovery of metabolite discrepancies associated with disease rather than differential urine dilution. Further work investigated the effect of identified xenobiotics in urine samples on the discrimination between individual subjects. This revealed that in these healthy patients, the detected xenobiotics had a minimal effect on discrimination between subjects compared to that of fluctuations of endogenously derived metabolites.

4.2 Introduction

As with any field of study, the presence of bias threatens to undermine the analysis of data and its subsequent interpretation. Within the field of urinary metabolomics this is no different; in fact due to the nature of urine there is a unique set of confounding factors. One of the two factors with the most potential to skew statistical analysis is varying urine concentration, and this has been extensively discussed in the literature with no consensus on the optimal normalisation approach (Warrack et al., 2009, Chen et al., 2013b, Veselkov et al., 2011a, Heavner et al., 2006, Mattarucchi and Guillou, 2012, Mattarucchi and Guillou, 2011, Jacob et al., 2014b, Edmands et al., 2014). The other is the presence of xenobiotics, either from the diet or pharmaceutical intake, and these have the potential to mask variation in the

endogenous metabolome (Walsh et al., 2006, Fernández-Peralbo and Luque de Castro, 2012, Lenz et al., 2004).

A major factor contributing to metabolome variation is urine concentration. Unlike most other biofluids, the volume of urine is not physiologically controlled. Instead it is directly linked to the individual's hydration status, primarily through water consumption (Warrack et al., 2009, Chen et al., 2013b, Veselkov et al., 2011a). As such, urine volume has been observed to vary by up to 15 fold under normal healthy physiological conditions, thus resulting in a large variation of metabolite concentration through dilution/concentration (Chen et al., 2013b, Veselkov et al., 2011a). To counter this, the normalisation of urine metabolite abundance is required to minimise the effect of metabolite dilution/concentration as a result of variable urine volume (Heavner et al., 2006, Warrack et al., 2009, Chen et al., 2013b, Veselkov et al., 2011a).

There are two stages in the experimental process where normalisation can be performed, pre- and post-analysis. Pre-acquisition normalisation acts as a preventative normalisation method, whereas post-analysis normalisation has been referred to as curative technique (Edmands et al., 2014). To date, the majority of normalisation techniques have focussed on the curative methods and require the acquired peak intensities to be divided by a normalisation factor.

4.2.1 Post-acquisition normalisation methods

Until recently, post-acquisition normalisation dominated urine metabolomic studies. In these studies the peak area for each metabolite in a chromatogram was divided by a normalisation factor such as; urine creatinine (Heavner et al., 2006), specific gravity (Heavner et al., 2006), osmolality (Boudonck et al., 2009), urine volume (Godzien et al., 2013) and several purely mathematical methodologies (Veselkov et al., 2011a).

Many initial urine metabolomics studies used a clinical measurement of urine concentration, urinary creatinine, to normalise peak intensities. Creatinine is a normal endogenous metabolite present in human urine and is a by-product of the breakdown of creatine phosphate in muscle tissue. Creatinine has been extensively used as a surrogate marker of urine concentration, and is still used to standardise concentrations of individual metabolites in urine (Warrack et al., 2009, Heavner et al., 2006, Hayashi et al., 2011, Saude et al., 2007, Bouatra et al., 2013, Zamora-Ros et al., 2011, Sterz et al., 2012, Schnackenberg et al., 2012, Ganti et al., 2012a). This normalisation methodology makes the assumption that creatinine excretion is constant for each individual. However, creatinine concentrations have

been observed to vary by up to 5 fold as a result of a number of factors including diet, activity, gender and health (Warrack et al., 2009, Heavner et al., 2006). In addition, a urine metabolomic study of male and female rats on different dosing regimens by Warrack et al., (2009) demonstrated that post-analysis normalisation to creatinine still discriminated based on sample dilution in the same manner to that of a non-normalised data set.

Due to the disparity in individual urine volumes between voids it has been suggested that the total volume of the void may be used as a normalisation technique (Godzien et al., 2013). However, using this approach Warrack et al., (2009) was only able to discriminate between two different dose groups (high and low), but not between male and female samples. Furthermore, in a later study normalisation to urine volume performed worse than post-analysis normalisation to specific gravity, however no explanation for this was provided (Edmands et al., 2014).

As an alternative approach, osmolality has come to the forefront of normalisation and is considered the gold standard for determining urine concentration (Chadha et al., 2001). The osmolality of urine is typically ascertained by freezing point depression (Chadha et al., 2001) and is as a measure of endogenous metabolic excretion (Warrack et al., 2009). Unlike other methods such as creatinine and specific gravity, osmolality is not affected by proteinuric samples or the presence of other high molecular weight compounds such as contrast agent (Chadha et al., 2001). Indeed Warrack et al., (2009) demonstrated osmolality normalisation to be capable of discriminating between treatment groups and gender, in addition to reducing variation in biological replicates. As such, osmolality normalisation has been adopted for a number of recent urine metabolomics studies (Boudonck et al., 2009, Warrack et al., 2009, Kim et al., 2011, Wittmann et al., 2014).

The use of specific gravity has also become a common method to determine urine concentration. Specific gravity is the ratio between the density of urine and pure water. It can be measured using refractometry, gravitometry or reagent strips (Chadha et al., 2001, Edmands et al., 2014). The specific gravity of urine may then be used as a normalisation factor for metabolomic analysis. However, specific gravity measurements may be skewed due to protein and other high molecular weight compounds such as contrast agents. These compounds disproportionately affect both refractometry and gravitometry measurements (Chadha et al., 2001). Despite this, specific gravity normalisation has been used successfully in urine metabolomics when measured using refractometry (Heavner et al., 2006, Jacob et al., 2014b, Edmands et al., 2014, Jacob et al., 2014a). While osmolality is considered the gold

standard for determining urine concentration, a strong positive linear relationship between osmolality and specific gravity has been observed, suggesting that both measurements can be used interchangeably to determine urine concentration (Chadha et al., 2001).

4.2.2 Mathematical post acquisition normalisation

So far the methods discussed have relied on various properties of urine and using these values to normalise the dataset. There are however, several mathematical post-analytical techniques which are widely implemented for the normalisation of urine metabolomic data. The two most commonly utilised are mass spectrum total useful signal (MSTUS) and mass spectrum total signal (MSTS). MSTUS calculates the sum peak intensity of metabolites within a chromatogram which are common to all samples, and uses this as a normalisation factor (Warrack et al., 2009, Mattarucchi and Guillou, 2012). All peak intensities within a chromatogram are subsequently divided by the chromatograms MSTUS normalisation factor. This means that the sum peak intensity for all common peaks regardless of the sample are equalised. The more of the common factors detected, then the more representative the normalisation factor is (Mattarucchi and Guillou, 2012). Confounding factors such as pharmaceuticals and dietary factors which are not suitable for calculation of a normalisation factor are unlikely to be present in all individuals and thus will not be utilised (Warrack et al., 2009). Warrack et al., (2009) found MSTUS to normalise urine metabolomic data to the same extent as osmolality, resulting in the differentiation between dose groups and gender in addition to reducing variation between biological replicates. The equation for MSTUS is given below in Eq 4.1.

$$MSTUS = \frac{Peak\ intensity\ of\ any\ peak\ in\ given\ chromatogram}{Sum\ peak\ intensity\ of\ all\ common\ peaks\ in\ the\ same\ chromatogram}$$

Eq.4.1 Calculation for MSTUS.

(Common peaks refer to peaks common to all samples in the study dataset).

MSTS differs from MSTUS in that it uses the sum peak intensity for all peaks within a chromatogram as the sample normalisation factor, and this means that after normalisation the sum peak intensity of all chromatograms will be equal (Ganti and Weiss, 2011). This methodology is built into several software packages used for chromatogram peak picking and deconvolution such as Waters MarkerLynx and Waters Progenesis (Waters, Manchester, UK) and is an extensively utilized form of normalisation (Ganti and Weiss, 2011, Kind et al., 2007, Beger et al., 2008, Manna et al., 2011, Banday et al., 2011, Zhang et al., 2010). The equation (Eq. 4.2) below shows the calculation for MSTS normalisation.

$$MSTS = \frac{Peak\ intensity\ of\ any\ peak\ in\ given\ chromatogram}{Sum\ peak\ intensity\ of\ all\ peaks\ in\ the\ same\ chromatogram}$$

Eq.4.2 Calculation for MSTS

Other mathematical normalisation methods have been investigated by Vesselkov et al., (2011a). Here four normalisation methods were analysed; MSTS, peak area median fold change, locally weighted scatter plot smoothing and quantile normalisations. The MSTS methodology was the same as described previously. Peak area median fold change normalises the log fold changes in peak intensities between samples. Once normalised the median log fold change is close to zero. Quantile normalisation ensures that all samples in the batch have an identical peak intensity distribution (Mar et al., 2009). This method may potentially be skewed by very high intensity peaks which vary between samples, or are not present in all samples. These confounding metabolites could include pharmaceuticals, making this normalisation technique undesirable in clinical studies. Locally weighted scatter plot smoothing forces the local median peak intensity fold change to be zero. This method is peak intensity dependent and assumes that different scaling factors are required for different peak intensity ranges (Veselkov et al., 2011a). In this study, all methods performed well for correcting for serially diluted urine samples. However, normalisation to median fold change was found to be preferential in the presence of biological variation (Veselkov et al., 2011a). These methods however, are not as widespread as some of the other previously described methods and they also require knowledge of R and use of command line programming methods.

The limiting factor with these correctional normalisation methodologies is that they are applied following LC-MS analysis. This means normalisation is poor for metabolites that do not follow a linear dilution pattern. These include low abundance metabolites that are diluted below the limit of detection in more hydrated samples, or concentrated metabolites that exceed the dynamic range of the mass spectrometer thus leading to detector saturation (Chen et al., 2013b, Mattarucchi and Guillou, 2011).

4.2.3 Pre-acquisition normalisation methods

To avoid problems that arise due to non-linear responses of the mass spectrometer, or concentration-dependent formation of artefacts such as dimer and adducts, the injection volume or urine concentration can be normalised to a physical property of urine prior to LC-MS analysis as a preventative normalisation method.

4.2.3.1 Injection volume correction

One method to correct for urine concentration prior to analysis is to vary the injection volume according to urine concentration. One such study has investigated this methodology by adjusting the injection volume to reflect urine creatinine levels (Chen et al., 2013b). This method offered improved normalisation for serially diluted rat urine when compared to post-analysis normalisation to either creatinine concentration, mass spectral creatinine peak area, MSTS or MSTUS. Further improvement on this pre-analysis normalisation was observed with the addition of a post-analysis MSTUS normalisation, which reduced variation in mass spectrometer response (Chen et al., 2013b). Nonetheless, this technique may suffer from differing chromatographic dilution due to varying sample injection volumes. This allows compounds to elute from the analytical column at different concentrations between samples. This effect will be magnified on nLC platforms where injections are into flow rates of <1000nl/min.

4.2.3.2 Urine concentration correction by dilution

The undesirable chromatographic dilution caused by varying injection volume can be avoided by equalising urine concentration prior to sample injection. This can be achieved by diluting or concentrating urine samples to give a uniform urine concentration. Using techniques that achieve this will allow equal concentrations of urine to be analysed by LC-MS.

In one study by Jacob et al., (2014b), the effect of pre-analysis freeze drying, and normalisation to specific gravity and post-analysis MSTUS normalisation techniques were investigated. The freeze dried urine samples were reconstituted in a volume reflecting the dried weight of urine (30mg/mL), to ensure the same weight of urine was analysed. Whereas, those samples normalised to specific gravity were diluted down to the urine with the lowest specific gravity. In this study, it was found that the specific gravity and MSTUS methodologies performed in a similar manner, and both improved differentiation between sample groups compared to the freeze drying normalised samples. This is potentially due to metabolites being lost during the freeze drying process and residual protein and salt levels skewing the dry weight of urine. A further advantage of both MSTUS and specific gravity normalisation is that it greatly increases sample through-put compared with the time consuming freeze drying process (Jacob et al., 2014b). The use of specific gravity pre-analysis normalisation has subsequently been adopted for future studies by Jacob et al. (Jacob et al., 2015, Jacob et al., 2014a).

A further study investigating specific gravity normalisation quantified the differences between pre- and post-analysis normalisation. Compared to non-normalised data sets, post-analysis normalisation to specific gravity led to a 2.3 fold increase in discriminant MS markers (Edmands et al., 2014). This difference was increased to 4.2 fold by pre-analysis normalisation via the reconstitution of samples in a specific gravity scaled volume (Edmands et al., 2014). This result confirms that normalising urine concentration prior to LC-MS analysis results in a more comprehensive normalisation method compared to post-analysis normalisation methods.

In 2012, Mattarucchi and Guillou demonstrated a novel pre-normalisation method using MSTUS which significantly improved the normalisation of urine samples. This methodology however requires samples to be analysed twice. Once to estimate the urine concentration using MSTUS and then again to analyse the MSTUS pre-normalised urine samples. This however, significantly increased the quantity of sample volume consumed and the time taken for analysis (Mattarucchi and Guillou, 2012). However, interestingly MSTUS normalisation factors and osmolality revealed a direct linear relationship, which suggests that that MSTUS, specific gravity or urine osmolality may be used interchangeably (Warrack et al., 2009).

4.2.4 Xenobiotic effect

Due to urine being one of the major routes of excretion of xenobiotics and their metabolites, they can be found in high abundance in urine (Holmes et al., 2007, Johnson et al., 2012a). Xenobiotics can be detected as unchanged parent molecules or a range of different metabolites derived from phase 1 and/or phase 2 metabolism in tissues, including the liver, and in some cases further metabolism by the gut microflora (Kell and Goodacre, 2014, Medina et al., 2013a). There have been a number of studies investigating large cohorts where cultural differences can be elucidated based upon dietary metabolites (Lenz et al., 2004, Holmes et al., 2008). Indeed in one study, a sample was noted as an outlier due to the individual being on the Atkins diet and thus displayed unusually high levels of taurine, while others were shown to differ due to their pharmaceutical intake (Lenz et al., 2004). In an investigation into acute kidney injury (AKI) as a result of cardiothoracic surgery, one of the discriminating metabolites identified was a pharmaceutical administered to the patients, rather than a underlying endogenous metabolite (Zacharias et al., 2013). The differences in pharmaceutical intake in clinical metabolomic studies are potentially a large source of discrimination between patient populations which may mask under lying metabolomic variation due to disease. It is estimated that there are over 25,000 dietary metabolites which may consumed in an individual's diet,

and potentially excreted in the urine (Scalbert et al., 2014). This is of particular concern to the discovery of biomarkers of disease or toxicity, as dietary differences may enhance the metabolomics variability of samples and thus hinder the discovery of relevant biomarkers. These issues have been highlighted as a potential issue in urinary metabolomics, but as yet has received little attention (Fernández-Peralbo and Luque de Castro, 2012).

In animal studies it is relatively simple to control the exposure to xenobiotics by controlling the diet and other sources of xenobiotics such as pharmaceuticals. However, in human clinical trials this is not possible and is potentially unethical (Gu et al., 2007, Lenz et al., 2004). It has been demonstrated in rats that changing diets at 24 hour intervals leads to significant changes in the urinary metabolome (Gu et al., 2007). Two human studies reduced the variation in urinary metabolome profiles by implementing a standardised diet either on the day prior to, or on the day of, urine collection (Walsh et al., 2006), (Lenz et al., 2003). However, this is not feasible for clinical metabolomic studies and would not exclude for pharmaceutical intake.

4.2.5. Study aims

The aim of this study was to establish the most appropriate method for the normalisation of human urine concentration. To achieve this, 5 urine samples each with 3 dilutions were analysed using LC-MS. The acquired dataset were initially normalised by post-analysis to the MSTS, or to creatinine or osmolality concentrations. Subsequent analysis used pre-analysis normalisation to osmolality, where all samples were diluted down to the osmolality of the most dilute sample. This ensured all urine samples had the same osmolality prior to analysis. Osmolality was chosen for the pre-analysis normalisation technique due to its direct proportional relationship to specific gravity which has already been shown to be a robust pre-normalisation method. In addition, the measurement of urine osmolality is a standard test in a clinical laboratory thus making it ideal for clinical case studies. Once the optimal normalisation methodology was determined, the effect of excluding detected xenobiotics on the classification of sample datasets was investigated.

4.3 Materials and methods

4.3.1 Chemicals

HPLC grade solvents were purchased from Rathburn Chemicals Ltd (Walkerburn, Scotland, UK) and UHPLC grade solvents from Fisher Scientific UK (Loughborough, UK). Strata X-AW 60 mg/3ml solid phase extraction (SPE) cartridges were purchased from Phenomenex (Macclesfield, U.K). Deuterated internal standards (IS); 17β-estradiol 2,4,16,16-d4 sodium 3-

sulfate (E2-d4-S, >99% D atom), carbamazepine (ring-d10), venlafaxine (N,N-dimethyl-d6), and diclofenac (phenyl-d4) were purchased from Cambridge Isotope Laboratories Inc. (MA, USA). Progesterone-2,2,4,6,6,17R,21,21,21-d9 (P-d9, 98% D atom) was purchased from CDN isotopes (Quebec, Canada) and prostaglandin E2-d4 (9-oxo-11α,15S-dihydroxy-prosta-5Z,13E-dien-1-oic-3,3,4,4-d4 acid) was purchased from Cayman Chemical Company (MI, USA). All other standards and reagents were purchased from Sigma-Aldrich Company Ltd., Dorset, U.K. All solvents used were HPLC grade with >99% purity.

4.3.2 Sample collection

Urine samples were collected from overnight fasting male patients in an HIV clinic at Brighton and Sussex University Hospital. The samples chosen were from HIV/hepatitis negative patients and were immediately frozen at -80 °C. All patients provided written, informed consent to the NHS Research Ethics Committee (NREC 09/H1107/101).

4.3.3 Sample preparation

Urine samples were defrosted and vortex mixed. Aliquots of 1.5 mL from 5 HIV negative individuals were taken. This aliquot was used to produce neat, 1 in 2 and 1 in 5 dilutions using water to give a 700 µL sample at each concentration. Samples were spiked with 0.5 ng/µL IS and centrifuged at 13000 rpm for 10 minutes. The resulting supernatants were adjusted to pH 2 with formic acid and extracted using mixed mode weak anion exchange and strong cation exchange solid phase extraction cartridges. Due to only 1.5 mL of each sample being available it was only possible to extract each dilution once. Strata X-AW and X-C cartridges were stacked with the X-AW above the X-C, both were conditioned with 1 mL methanol and 1 mL water. The X-AW cartridge was loaded with 1 mL of the acidified urine and washed with 1 mL water. Extracts were eluted with 1 mL 5% ammonium hydroxide in methanol followed by 1 mL ethyl acetate. The two solvent elutions were combined, the solvents removed under vacuum and the extracts reconstituted in 70 µL of 90: 10 water: methanol. Each 70 µL sample was split into two 35 µL aliquots. One of the 35 µL aliquots of each sample was diluted down to reflect a uniform (pre-SPE) osmolality of 44.5 mOsm, which is the lowest osmolality concentration measured in pre-extracted samples (that of the 1 in 5 dilution of individual sample number 440). Aliquot dilution was achieved using water spiked with 0.5 ng/µL IS thus ensuring the concentration of internal standard was uniform in all samples and for these samples this ensured that the same amount of urine (in terms of osmolality) was injected on column in every analysis. All samples were frozen at -80 ^{o}C prior to nLC-nESI-TOFMS analysis. In order to have enough data for analysis each dilution was injected 3 times as a LC-MS replicate.

Quality control samples (QC) were made up of 100 µL aliquots of each urine sample to give 500 µL of pooled urine. This was extracted using the SPE methodology described and reconstituted in 50 µL to give a 2 fold rather than a 10 fold concentration to reduce source fouling. A QC was injected 3 times before each batch and after every 10th sample within batches to assess MS performance throughout the run.

4.3.4 nUHPLC-nESI-TOFMS analysis

Samples (0.5 µL) were injected in triplicate in random order onto a Waters nanoAcquity and separated using a Waters nanoAcquity HSS-T3 (100mm x 100 µm x 2.8 µm, 100 Å) column. Chromatographic separation was carried out at 700 nL/min using UHPLC grade water and acetonitrile as mobile phases A and B respectively, both modified with 0.01% formic acid. A gradient elution was used: 0 min 10% B, 4 min 30% B, 18 min 50%B, 30 min 100%B, 100% B maintained for 10 min then equilibrated in initial conditions for a further 15 min. Metabolites were detected in positive ESI using a Waters Xevo G2 TOFMS equipped with a nano ESI source. The TOFMS was tuned to a mass resolution of 20,000.

4.3.5 Data analysis

The MS datasets were deisotoped, mass spectral peaks deconvoluted, aligned, and the datasets binned using Waters MarkerLynx software. Extracted markers were exported to Simca v13.0 (Umetrics Ltd, Crewe, UK) for multivariate analyses. All data were log transformed and Pareto scaled prior to principal components analysis (PCA) to identify the effect of different normalisation methods. Further modelling using orthogonal partial least-squares discriminant analysis (OPLS-DA) was used to investigate metabolite differences between individuals (Wiklund et al., 2008). Metabolite identities were determined from their accurate mass, isotopic fit, and comparison of fragmentation data with authentic standards or with Metlin (Tautenhahn et al., 2012a), Human Metabolome Database (Wishart et al., 2007), Human Urine Metabolome Database (Bouatra et al., 2013) and MycompoundID (Li et al., 2013a) databases.

4.3.6 Quality control and reproducibility

The quality of metabolomic analyses was assessed using the method proposed by Want et al., whereby the coefficient of variation was calculated for the mean peak area of peaks present in 80% of the QC samples. Reliable metabolomic analysis is achieved when the %CV falls below 30% for more than 70% of peaks present in at least 80% of QC samples (Want et al., 2010). This procedure was also carried out to determine the effect of normalisation on

metabolite mean peak area reproducibility. In this instance, the coefficient of variation for mean peak area was calculated for peaks common to all of the individual samples.

4.3.7 Normalisation methodology

4.3.7.1 MSTS

Normalisation to mass spectrum total signal was performed using Microsoft Excel where all peaks within a chromatogram are normalised to give equal sum peaks areas for all samples see previous Eq. 4.2).

4.3.7.2 MSTUS

This normalisation factor was calculated by summing the peak areas for markers common to all samples (see Eq. 4.1) using Microsoft Excel. As a result, each sample had its own normalisation factor based only upon the total intensities of peaks which are also common to all other samples, thus eliminating any bias as a result of xenobiotic and random background noise. All peaks areas within a chromatogram were subsequently divided by the sample MSTUS normalisation factor.

4.3.7.3 Creatinine and osmolality

Creatinine levels were determined at Brighton and Sussex University Hospital via the colorimetric Jaffe reaction (Heinegård and Tiderström, 1973, Husdan and Rapoport, 1968). Osmolality of urine samples were determined by freezing point depression using an Osmometer Model 3320 (Advanced Instruments Inc.). The peak area of each detected ion was divided by that of the sample creatinine or osmolality measurement using Excel software.

4.3.7.4 Pre-analysis normalisation to osmolality

The choice of osmolality for the pre-analysis normalisation for urine concentration was due to osmolality having a strong positive correlation to specific gravity which has already been successfully used for urine pre-analysis normalisation (Jacob et al., 2014b, Edmands et al., 2014). In addition, osmolality is a standard test for urine in a clinical laboratory making it ideal for medical metabolomic studies as no additional measurement outside of normal clinical practice is required. As described in the above section, aliquots of SPE extracted samples were reconstituted in a volume to reflect a constant pre-extraction osmolality value of 44.5 mOsm based on the osmolality measurement of the 1 in 5 dilution of samples 440 prior to SPE. This dilution was achieved using HPLC grade water spiked with 0.5 ng/µL IS to ensure uniform concentration of IS throughout the analysis. This allowed the same injection volume of 0.5 µL to be used for each samples thus eliminating the effect of non-linear dilution. Data from these

analyses were assessed as pre-analysis osmolality adjusted urine concentration, and in addition were also combined with post-analysis MSTS or MSTUS normalisation strategies.

4.3.9 Mass exclusion of xenobiotics

Once the optimal normalisation method had been determined, xenobiotics discriminating between individual subjects were identified and excluded from multivariate analysis. Once excluded, PCA analysis was performed to determine whether removal of xenobiotics from metabolomics data sets resulted in less variation between the urinary metabolomes of different individuals.

4.4 Results and discussion

4.4.1 nUHPLC-nESI-TOFMS reproducibility

The reproducibility of the LC-MS metabolomic analysis was assessed from the analysis of the 6 QC samples within the batch analysis by calculating the %CV for mean peak area for peaks present in 80% (5 of 6) of the QC samples. This revealed that the %CV of mean peak area of 74% of peaks present in 80% of QC samples was below 30%. This degree of variation is within the suggested minimum limit of 70% of peaks present in 80% of samples returning a %CV of peak area of less than 30% (Want et al., 2010).

A further test of analytical reproducibility is a visual analysis of the PCA scores plot containing the QC samples. In reproducible datasets the QC group will cluster together, and any analytical variability arising from source fouling of the MS or LC retention time shift will disperse the QC data points on the score plot. The non-normalised data set including the QC samples are presented in Figure 4.1. Here the QC samples are quite disperse, reflecting some variation in LC-MS analysis. However typically a scores plot analysis is undertaken only following a normalisation methodology. An additional scores plot in Figure 4.2 where the data set is normalised post-analysis to MSTS illustrates very good grouping of the QC as well as the three analytical replicates of each urine dilution. This suggests that the dataset is highly reproducible and that MSTS normalisation corrects for small drifts in mass spectral signal.

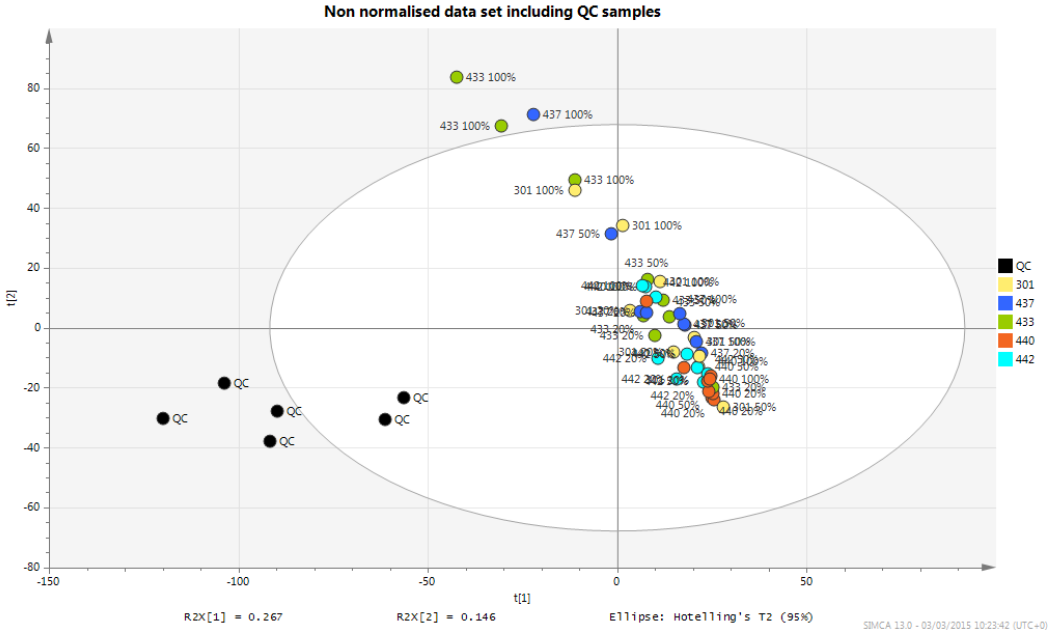


Figure 4.1: PCA scores plot analysis of the non-normalised data set of diluted urine and quality control samples

Legend shows samples coloured according to 5 different individuals. Sample dilution factor (100% no dilution, 20% 1 in 5 dilution, 50% 1 in 2 dilution) given on PCA plot. QC= quality control samples.

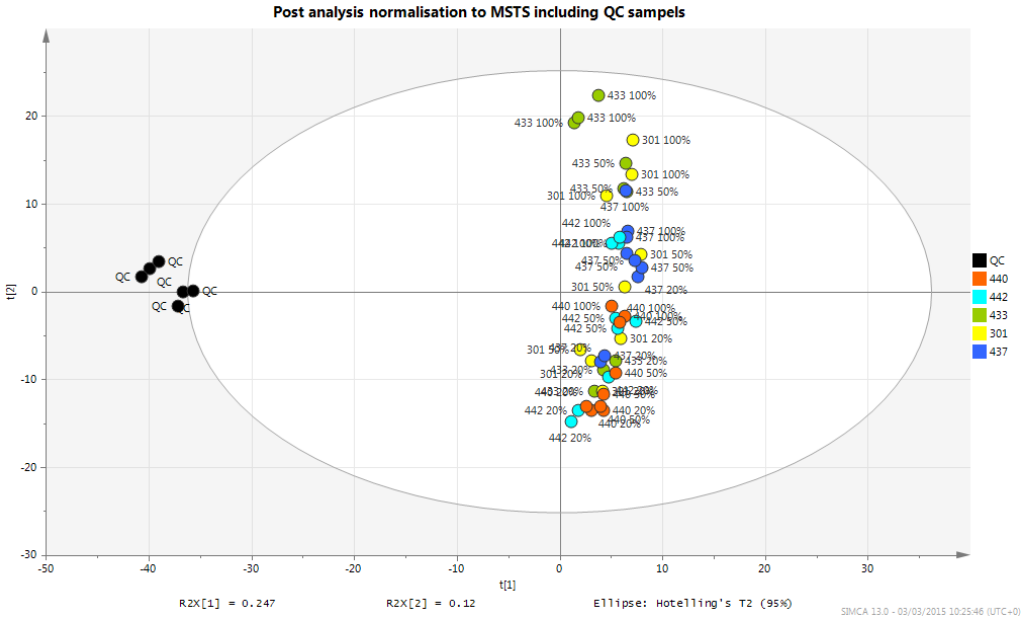


Figure 4.2: PCA scores plot analysis of data set from Figure 4.1 normalised post-analysis to MSTS

Legend shows samples coloured according to 5 different individuals. Sample dilution factor (100% no dilution, 20% 1 in 5 dilution, 50% 1 in 2 dilution) given on PCA plot. QC= quality control samples.

4.4.2 Urinary creatine and osmolality values

The concentrations of creatinine and osmolality measured in the samples for the 5 individuals are presented in Table 4.1. The osmolality values in the sample set varied from 44 to 1045 mOsm and reflected the range of values potentially present in a study population (see for example Table 6.1, Chapter 6). The creatinine and osmolality concentrations were used to normalise sample data sets in the following studies.

Table 4.1: Urinary creatinine concentrations and osmolality measurements

	Osmolality (mOsm)			Creatinine (mM)		
Sample	Neat	1in2	1in5	Neat	1in2*	1in5*
301	1045.5	532.7	228.3	19.2	9.6	3.8
433	908.0	460.5	182.3	35.1	17.6	7.0
437	691.5	345.5	141.5	7.4	3.7	1.5
440	212.5	106.5	44.5	4.7	2.4	0.9
442	614.3	311.5	121.5	10.7	5.4	2.1

*Creatinine concentrations for 1 in 2 and 1 in 5 dilutions were calculated from values measured in neat urine samples.

4.4.3 Normalisation of urine samples

As a result of the QC samples grouping separately from the samples of interest it was difficult to ascertain the effect of each normalisation technique on the remaining sample data set. As such, all further scores plots for the assessment of normalisation techniques have excluded the QC samples.

4.4.3.1 The effect of no post-analysis normalisation

The non-normalised data set is shown as a PCA scores plots in Figure 4.3. Here samples cluster on the scores plot as a result of dilution factor rather than by individual. This highlights the need for a normalisation methodology that corrects for urine dilution. Using a data set such as this would greatly hinder the identification of discriminating factors between individuals because the differential dilution of the urine drives the discrimination.

4.4.3.2 The effect of post-analysis normalisation to MSTS

The use of MSTS is a normalisation technique built into several software packages such as Waters MarkerLynx and Waters Progenesis and as such it has become a default normalisation method for many studies (Ganti and Weiss, 2011, Kind et al., 2007, Beger et al., 2008, Manna et al., 2011, Banday et al., 2011, Zhang et al., 2010). The PCA scores plot in Figure 4.4 highlights the problem faced with differential urine concentrations. Here discrimination is clearly driven by the concentration of urine on column with the 20%, 50% and 100% urine

concentrations clearly separated with minimal inter-individual discrimination. This is possibly a result of the signal of some ions being too low to be detected in the extracts of diluted samples thus influencing clustering on the PCA scores plot. Conversely, extracts of more concentrated neat samples may contain a greater proportion of adducts and di- or tri-mers in the sample data set resulting in these samples occupying a different chemical space to the diluted samples on the PCA scores plot.

4.4.3.3 The effect of post-analysis normalisation to MSTUS

The use of normalisation to MSTUS requires a subset of MS signals common to all samples in the dataset. As <5% of peaks were found to be common to all samples, which is likely due to the range of sample dilutions, then this technique could not be used for post-analysis normalisation of samples.

4.4.3.4 The effect of post-analysis normalisation to creatinine

The results of post-analysis normalisation to creatinine are displayed in the scores plot of Figure 4.5. Compared to the non-normalised data set, the sample grouping is very similar and driven by urine dilution. The data indicate creatinine concentration is a poor surrogate for measuring urine concentration of samples, possibly due to some of the reasons given previously (section 4.2.1) as its concentrations can vary markedly with diet and health. It also cannot correct for changes in detection limits due to loading of different urine concentrations on the LC-MS.

4.4.3.5 The effect of post-analysis normalisation to osmolality

The osmolality normalised scores plot in Figure 4.6 appears similar to that of the creatinine and the non-normalised data set. Discrimination is again driven by urine concentration but to a lesser extent than that seen in the MSTS normalisation method. There is no evidence that the use of osmolality post-analysis normalisation has improved either inter- or intra-individual clustering on the scores plot. Improvements in one or both of these factors would have suggested an improvement in the normalisation for urine concentration.

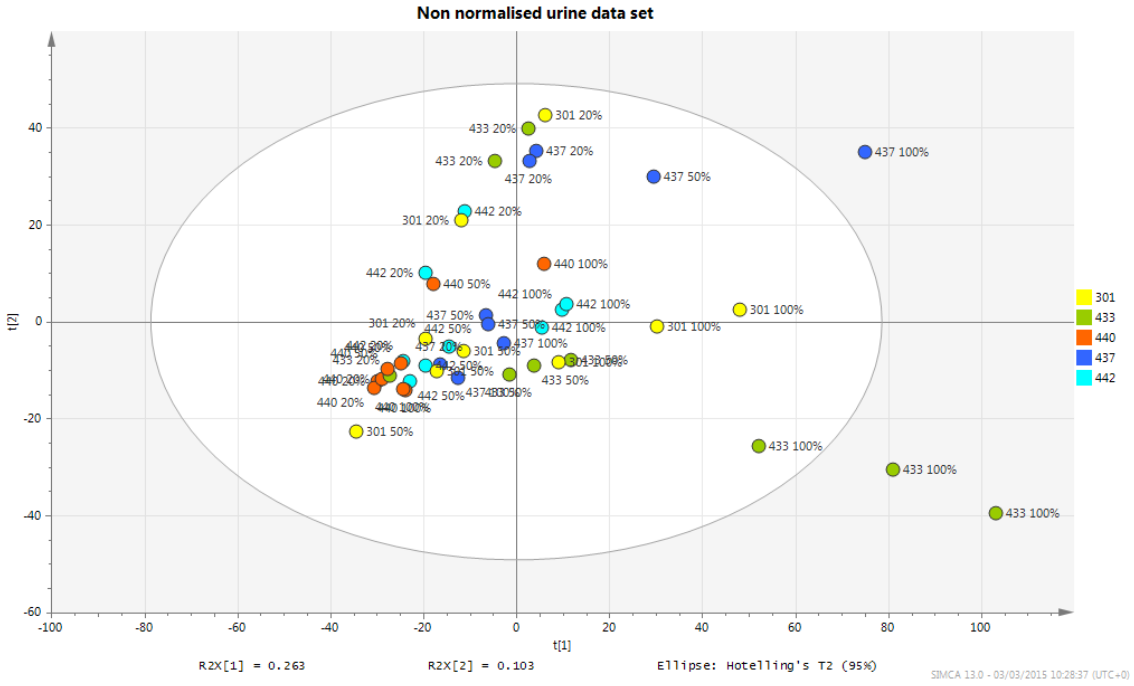


Figure 4.3: PCA analysis of the data set following LC-MS analysis of serially diluted urine samples without any normalisation

Legend shows samples coloured according to 5 different individuals. Sample dilution factor (100% no dilution, 20% 1 in 5 dilution, 50% 1 in 2 dilution) given on PCA plot. Without a correction for urine dilution multivariate analysis discriminates samples by dilution instead of discreetly grouping sample from the same individual.

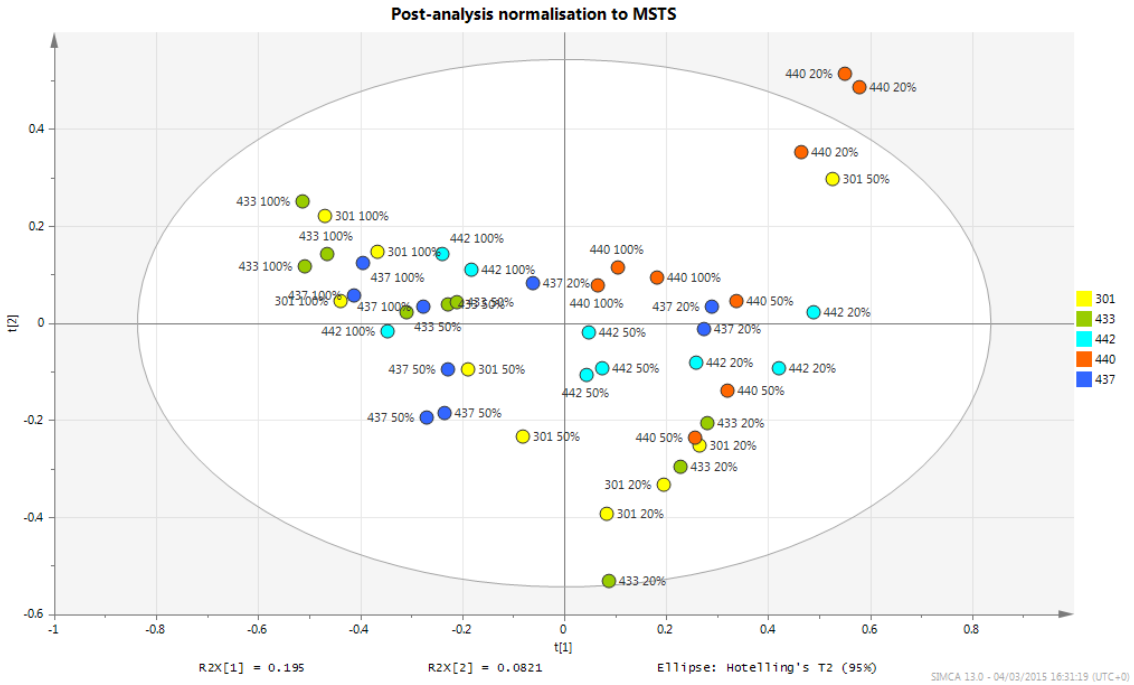


Figure 4.4: The effect of post-analysis normalisation to MSTS following LC-MS analysis of serially diluted urine samples

Legend as Figure 4.3. The scores plot discriminates based solely upon sample dilution factor with no inter-individual discrimination. The more concentrated samples are discriminated from the dilute samples by the first component.

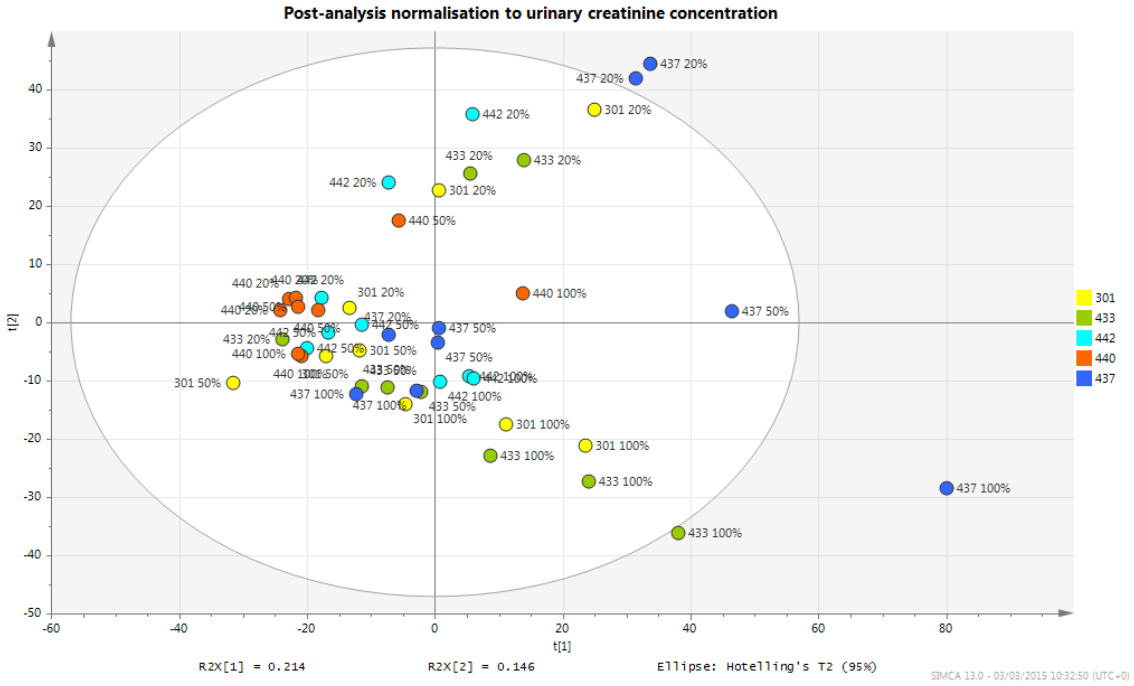


Figure 4.5: PCA scores plot of post-analysis normalisation to urinary creatinine concentration following LC-MS analysis

Legend as Figure 4.3. The scores plot is similar to that of the non-normalised data set. Discrimination is driven by the dilution factor of the urine rather than by each individual.

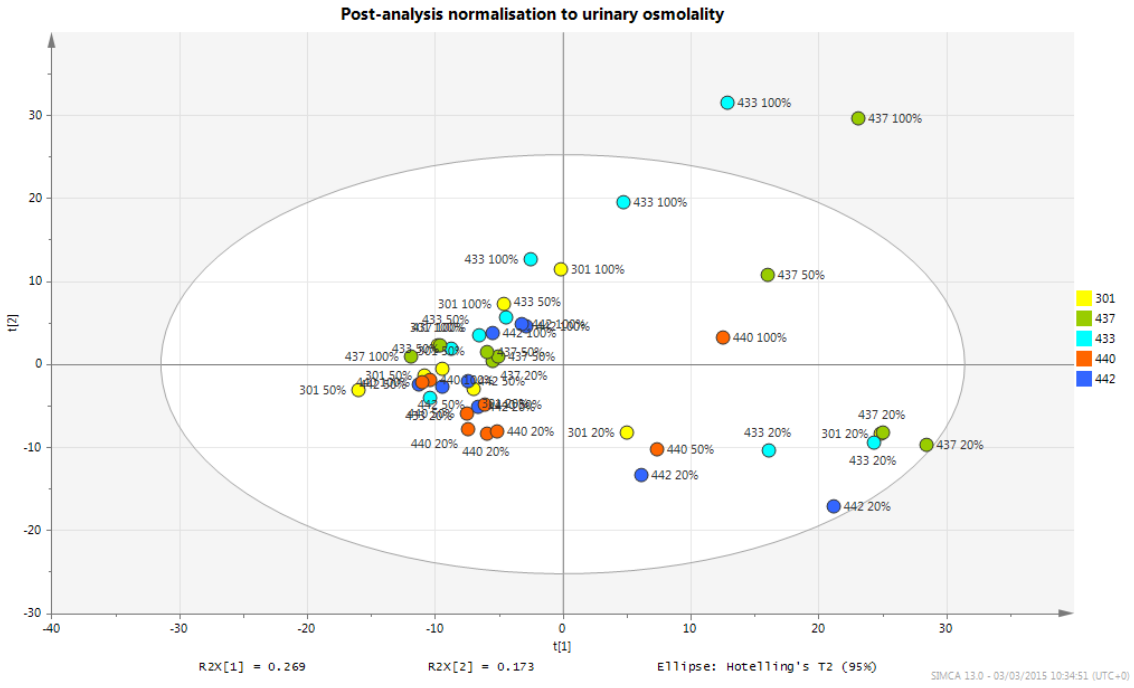


Figure 4.6: PCA scores plot of post-analysis normalisation to urinary osmolality following LC-MS analysis

Legend as Figure 4.3. The post-analysis normalisation is driven by the dilution factor of the urine sample. There is no indication of improvements in intra- or inter-individual clustering compared with no normalisation of the data set.

4.4.4 Pre-analysis normalisation

The results of the post-analysis normalisation methods highlight an important factor often over looked in urine normalisation in metabolomics. The majority of studies dilute their samples in water to eliminate the effect of detector saturation. This however, will exacerbate the loss of low abundance metabolites. The loss of compounds close to the limit of detection cannot be corrected for by post-analysis normalisation, and this leads to zero values which confound multivariate analysis (Mattarucchi and Guillou, 2012). Some groups resolve this problem with zero value data by adding an offset to their data prior to analysis (Veselkov et al., 2011b, Veselkov et al., 2011a) but despite this, low abundance metabolites just below the limit of detection in more dilute urine are not corrected for. Instead a method to inject the same amount of urine on column is needed. Despite there being very little visual difference in clustering on the scores plots for either post-analysis normalisation to creatinine or osmolality, osmolality was chosen as the pre-analysis normalisation factor. This is due to the reported effects of diet, gender, exercise and health on variations in creatinine concentration (Warrack et al., 2009) and the fact that osmolality is now considered the gold standard for determination of urine concentration. It also has a strong positive correlation to specific gravity which has been successfully implemented as a pre-analysis normalisation technique (Edmands et al., 2014, Jacob et al., 2014b). Osmolality was preferred to a MSTS or MSTUS pre-analysis normalisation technique as it allows a more high-throughput analysis without the need to analyse samples twice. As such, all urine samples were reconstituted in a volume to give an estimated concentration of 445 mOsm after SPE.

4.4.4.1 The effect of pre-analysis normalisation to osmolality

The analysis of osmolality-adjusted urine improved the within individual sample clustering on PCA scores plots (Figure 4.7). This is potentially a result of analysing the same amount of urine for every sample thus eliminating or at least equalising the effect of detector saturation or non-liner dilution effects on LC-MS analysis (Edmands et al., 2014, Chen et al., 2013b, Jacob et al., 2014b, Mattarucchi and Guillou, 2012). In addition, in some subjects discrimination on the scores plot was driven by the individual and not the dilution factor. It may be possible to further improve individual clustering with an additional post-analysis normalisation to reduce variability of the LC-MS analysis. The effect of MSTS has already been observed for reducing LC-MS variation and could potentially have the same result following a pre-analysis normalisation to osmolality. Furthermore, the addition of a post-analysis MSTUS normalisation has been shown elsewhere to improve upon pre-normalisation factors (Chen et al., 2013b).

4.4.4.2 The effect of pre-analysis normalisation to osmolality plus post-analysis normalisation to MSTS or MSTUS

As previously mentioned, an additional post-analysis normalisation to reduce the effect of LC-MS variation may further improve the quality of the metabolomic data set. Due to their wide spread use both MSTS and MSTUS methods were investigated; score plots for these analysis are presented in Figures 4.8 and 4.9 respectively. The additional MSTS normalisation appears to have improved some of the grouping for individuals. However, individual 440 is still relatively dispersed compared to the other individuals and clusters according to dilution. This grouping however is further improved when MSTUS is implemented instead of the MSTS. Here all individuals tend to cluster with the effect of dilution now being greatly reduced though not completely eliminated. It was possible that the remaining discrimination as a result of dilution could be caused by different loading of the urine concentrations onto the SPE cartridge during the sample preparation step, resulting in differences in extraction efficiency of some components of the metabolome.

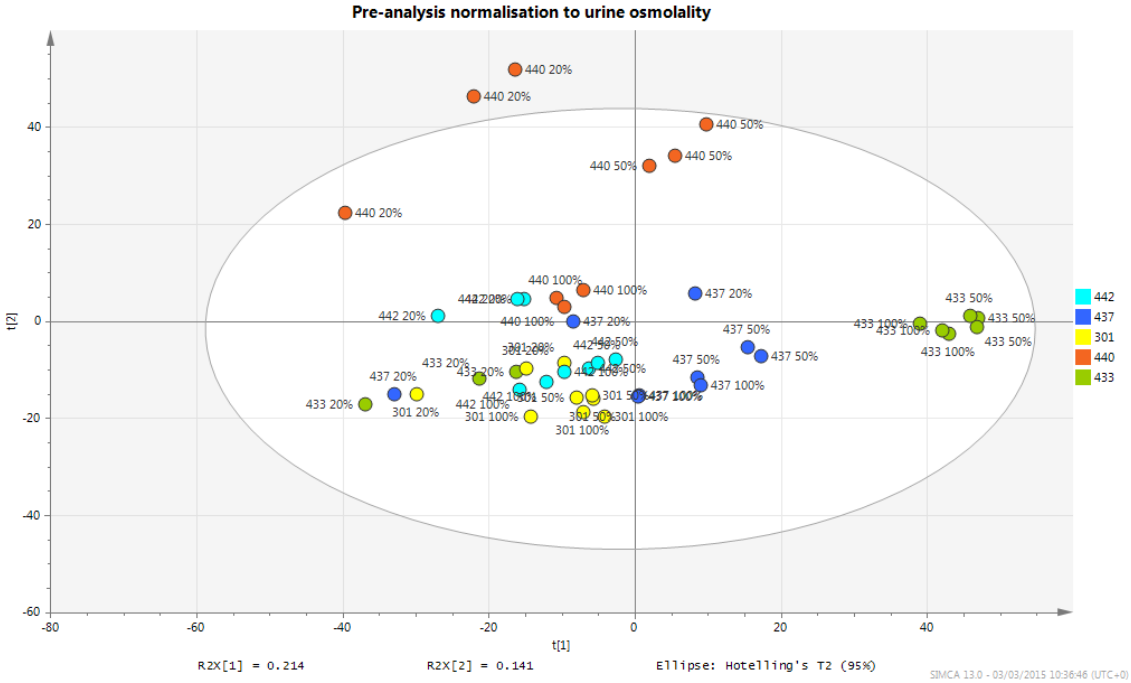


Figure 4.7: PCA scores plot of pre- analysis normalisation to osmolality prior to LC-MS analysis of serially diluted urine samples

Legend as Figure 4.3. The preventative normalisation to osmolality improves discrimination between individuals, with less variation within samples from the same individual.

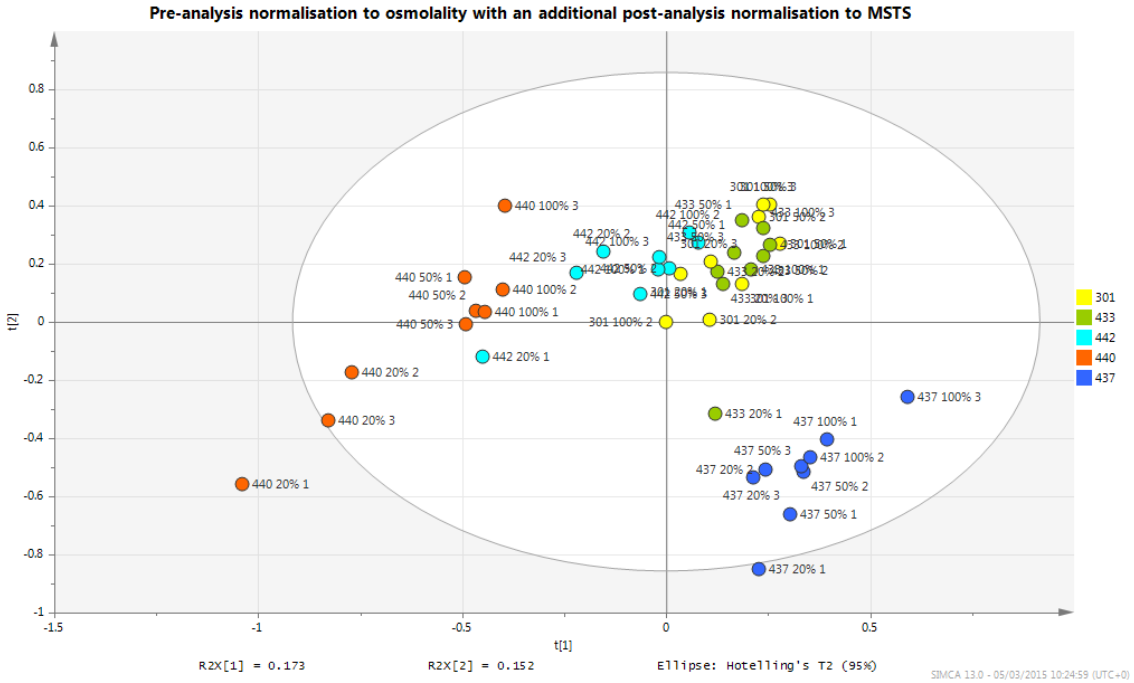


Figure 4.8: PCA scores plot of pre-analysis normalisation to osmolality followed by an additional post-analysis MSTS normalisation of serially diluted urine samples analysed by LC-MS

Legend as Figure 4.3. The addition of a further normalisation to MSTS post-analysis has improved intra-individual clustering on the scores plot with samples.

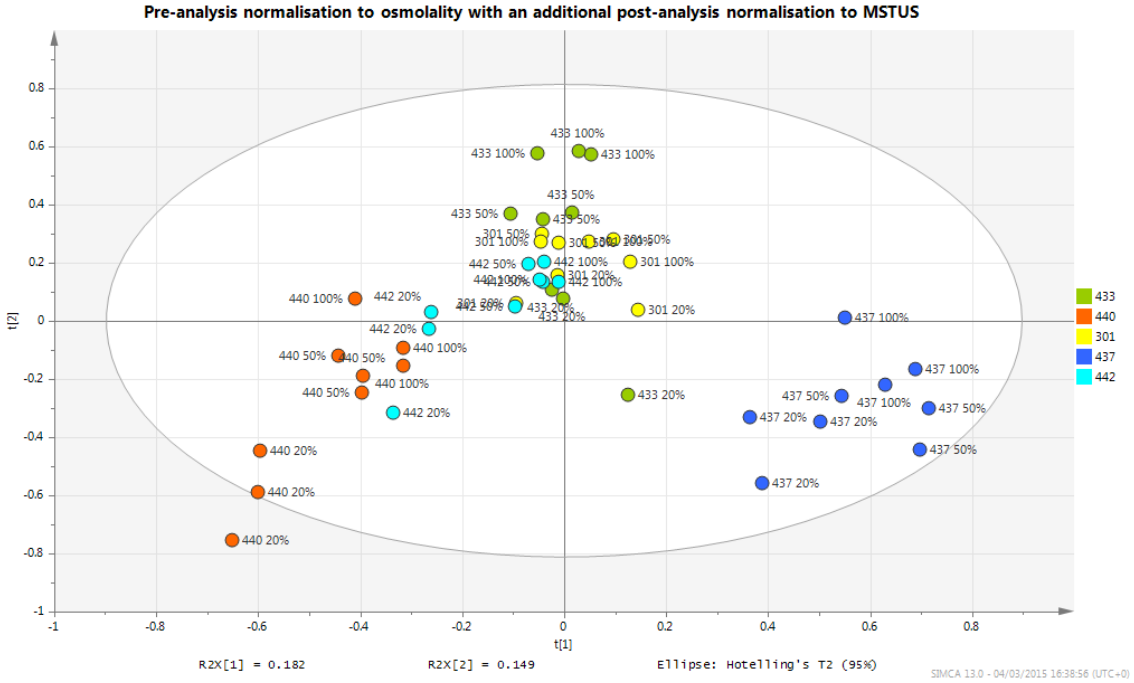


Figure 4.9: PCA scores plot of pre-analysis normalisation to osmolality followed by an additional post-analysis MSTUS normalisation of serially diluted urine samples analysed by LC-MS

Legend as Figure 4.3. In comparison to the pre-analysis to osmolality alone, grouping by individuals is more evident with less intra-individual variation of samples.

4.4.5 Quantitative analysis of normalisation methods

An additional method to determine the effect of differing normalisation techniques is to assess the mean peak area repeatability of metabolome (Jacob et al., 2014b). Using a slightly modified version of the method proposed by Want et al., (2010), the mean peak areas for peaks present in 100% of samples for each individual were determined and their %CV calculated. This analysis determines the total number of ions common to all dilutions of an individual and their respective mean peak area variation. An improved normalisation of the data set should result in an increase in common peaks detected and a greater percentage of these peaks will have a %CV <30%. These data are presented in Tables 4.2 and 4.3 respectively. These figures were only calculated for the non-normalised and post-analysis normalisation data sets only. This is because ≤6% of peaks were common within individuals in the non-normalised data set of which none have %CV of mean peak area below 30%. This helps to explain the poor post-analysis normalisation data sets because no post-analysis normalisation can correct for the loss of metabolites.

Table 4.2: The effect of normalisation on the number of peaks detected that were common to all samples

	Number of common peaks in each sample				
Normalisation method	**301**	**433**	**437**	**440**	**442**
Non-normalised	18	130	159	55	129
Pre-analysis to osmolality	557	546	598	641	614

Table 4.3: The effect of normalisation methodologies on mean peak area reproducibility

Normalisation method	**Sample**				
	301	**433**	**437**	**440**	**442**
Non-normalised	0%	0%	0%	0%	0%
Pre-osmolality	60%	20%	28%	26%	49%
Pre-osmolality + MSTS	66%	24%	18%	44%	56%
Pre-osmolality + MSTUS	65%	26%	25%	45%	57%

The percentages expressed are the percent of peaks common to each sample with a %CV of less than 30%

The use of pre-analysis normalisation to osmolality increased the number of commonly detected peaks by up to 31 fold, of which 20 to 60% returned a %CV for mean peak area below 30%. In contrast, none of the common peaks detected in the non-normalised data set had a mean peak area CV of <30%. These values reflect the improvements in clustering of samples from the same individual that were observed in the PCA scores plots using pre-analysis normalisation to osmolality method. The addition of a post-analysis MSTS or MSTUS normalisation to reduce variation in the LC-MS analysis generally increased the percentage of

peaks with a mean peak area %CV of <30% by a further 4 and 19% depending on the sample source (the sample from individual 437 being the exception). These data suggest that pre-analysis normalisation to osmolality with an additional post-analysis normalisation to MSTUS or MSTS offers an improved score plot clustering reducing variation associated with sample dilution in addition to providing a more repeatable metabolomic analysis. This methodology offers the best normalisation technique of those tested in this study, however some discrimination still remains as a result of urine dilution factor. This may be due to loading SPE cartridges with different urine concentrations. By diluting urine samples to equalise osmolality prior to SPE, further improvements in scores plot scattering may have been observed. An additional consideration is that in this study urine samples were diluted down to the lowest measured osmolality. This means that only the most dilute urine samples were analysed meaning that small variations in peak intensity have a greater impact on repeatability. There is also the added risk that important low abundance metabolites such as eicosanoids and unconjugated sex steroids were not detected as these dilutions. To counter this, future studies may benefit from reconstituting samples to a common osmolality chosen in the middle of the normal urine osmolality range of a population in order to reduce the risk of non-detection of low abundant metabolites.

4.4.6 Mass exclusion of xenobiotics

The effect of xenobiotics on multivariate analysis was investigated using OPLS-DA analysis to compare samples from individuals which clustered apart in the PCA model, i.e. individuals 433 and 437, and 433 and 440. Samples from these subjects were clearly separated on the PCA scores plot (Figure 4.9). Visualisation of discriminating chemicals between individuals 433 and 447 using an S-plot (see Figure 4.10) enabled the identification of 3 xenobiotics and 6 endogenous metabolites in which concentrations were markedly higher in individual 447 compared to 443 (see Table 4.4). The signal intensity plots for three of the compounds indicated that they were xenobiotics as they were clearly present in individual 437 but not detected at all in individual 433 (see Figure 4.11 for example). These compounds were identified as Ranitidine (a histamine H_2-receptor antagonist), a metabolite of acetaminophen, and a dietary polyphenol. The remaining 6 markers of individual 437 were endogenously derived metabolites, identified as carnitines and were also present at a lower concentration in individual 433 (see Figure 4.12 for example). Following the exclusion of the three xenobiotics a new PCA scores plot was constructed (Figure 4.13), which shows little change from the initial scores plot in Figure 4.9. This suggests that the discrimination seen on the PCA scores plot between individuals 433 and 437 was not driven solely by xenobiotics. To determine if this is

due to the 6 endogenous metabolites found to be driving separation these were also excluded (see Figure 4.14). The removal of all 9 (xeno)metabolites resulted in a redistribution of scores on the scores plot. Individual 437 has begun to group with that of 301 suggesting their metabolite profiles are more similar following the exclusion of the 9 (xeno)metabolites. Despite the exclusion of these (xeno)metabolites some discrimination remains between individuals 433 and 437. Further work examining discriminating compounds between individuals 433 and 440 revealed that they comprised over 40 compounds which were all endogenously derived metabolites. The data suggest that for the samples from these 5 individuals, that many metabolites are driving clustering of samples on the PCA model rather than a few xenobiotic compounds. However, the exclusion of xenobiotics in a clinical case study may have a much more significant effect on sample clustering in a PCA model, as patients may have been treated with high doses of a variety of pharmaceuticals related to their condition, e.g. in antiretroviral medication.

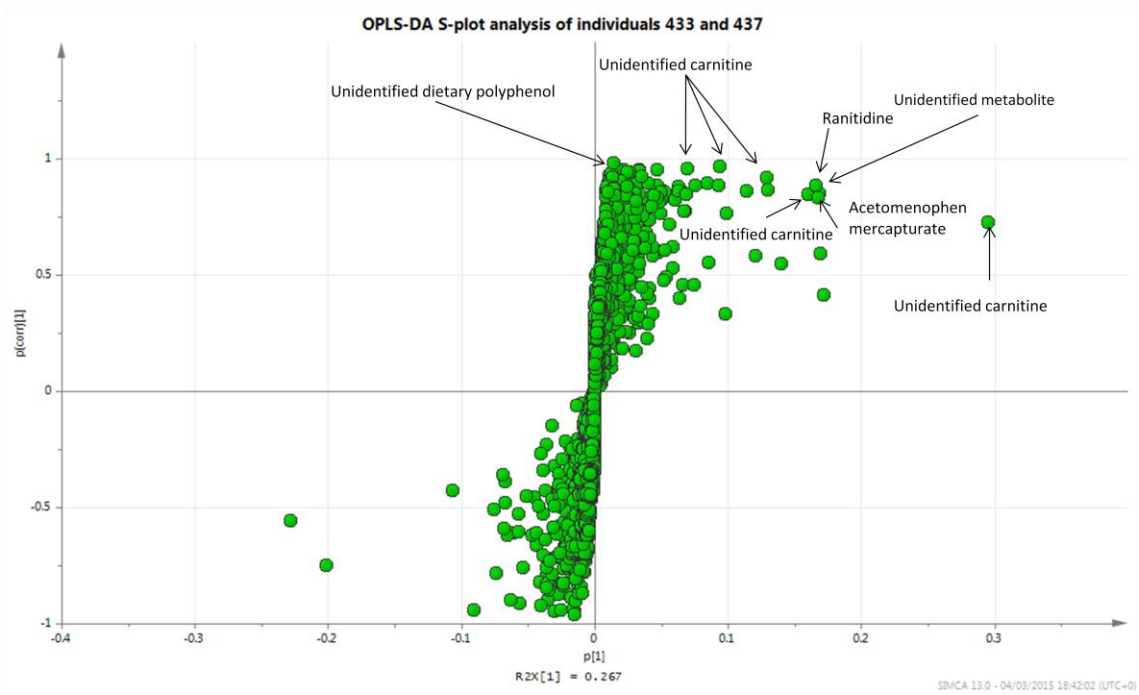


Figure 4.10: OPLS-DA S-plot analysis of discriminating metabolites between individuals 433 and 437

Metabolites found to be driving discrimination are labelled and removed from the PCA model.

Table 4.4: (Xeno)metabolites highlighted by OPLS-DA S-plot as discriminating between individuals 433 and 437

Experimental mass	Formula	Actual mass	Fragments	Identity
315.149	$C_{13}H_{22}N_4O_3S$	315.1491	176.0502	Ranitidine
271.076	$C_{11}H_{14}N_2O_4S$	271.0753	167.0575	Acetaminophen mercapturate
457.222	$C_{26}H_{32}O_7$	457.2226	None	Unidentified polyphenol
304.212	$C_{15}H_{29}NO_5$	304.212	85.0299	Unidentified carnitine
330.264	$C_{18}H_{35}NO_4$	330.2644	85.029	Unidentified carnitine
300.218	$C_{16}H_{29}NO_4$	300.2175	85.0297	Unidentified carnitine
356.243	$C_{19}H_{33}NO_5$	356.2437	85.0296	Unidentified carnitine
308.186	$C_{17}H_{26}NO_4$	308.1862	85.03	Unidentified carnitine
319.166	$C_{17}H_{22}N_2O_2$	319.1658	None	Unidentified metabolite

Concentrations of identified compounds were markedly higher in urine of 437 compared to 433.

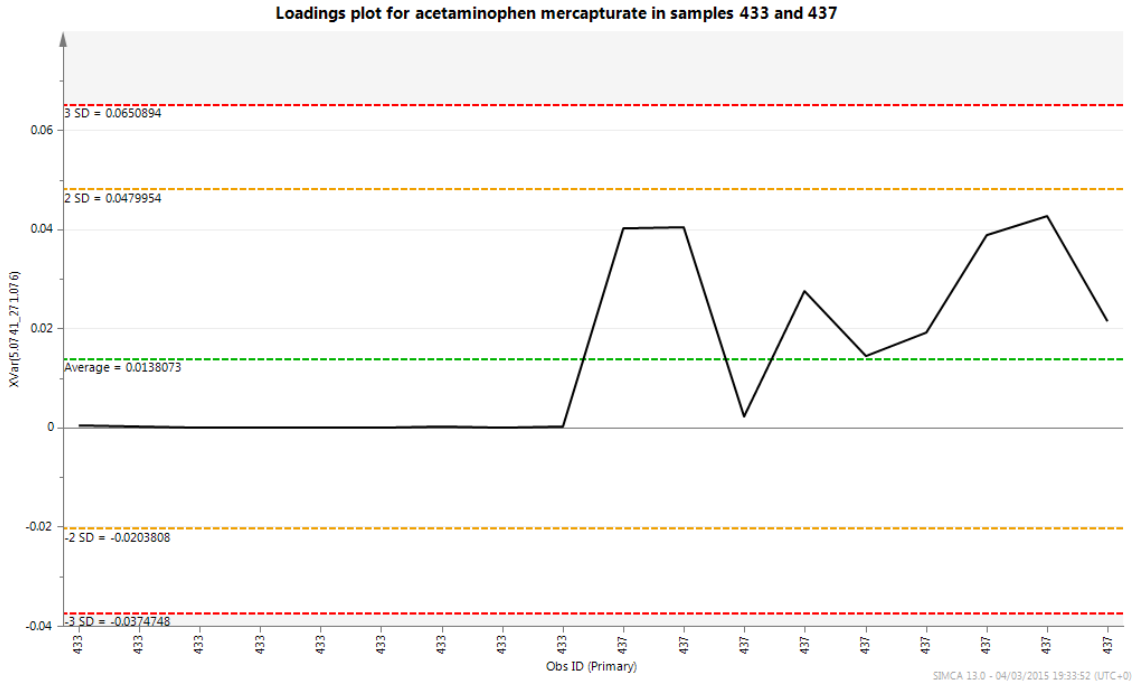


Figure 4.11: Signal intensity of acetaminophen mercapturate in samples 433 and 437

Acetaminophen mercapturate is present only in the urine sample of individual 437 thus driving some of the discrimination between the two individuals.

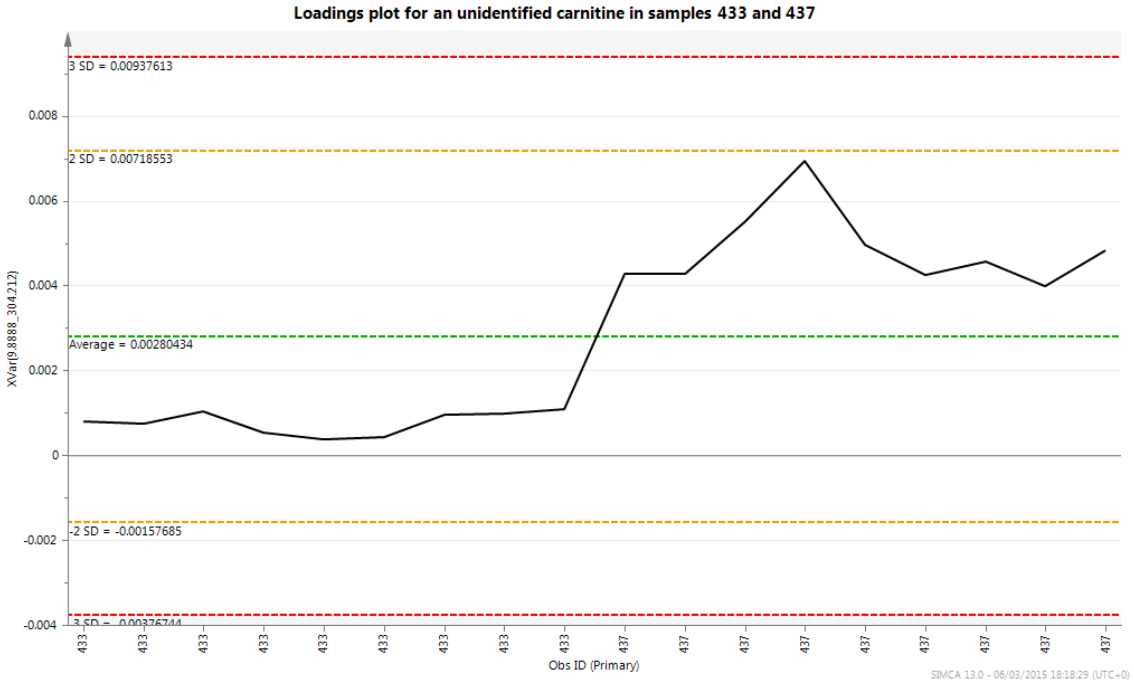


Figure 4.12: Signal intensity of an unidentified carnitine in samples 433 and 437

The unidentified carnitine is present in both individuals suggesting it is an endogenous metabolite. The elevated concentration in individual 437 drives some of the discrimination between sample groups.

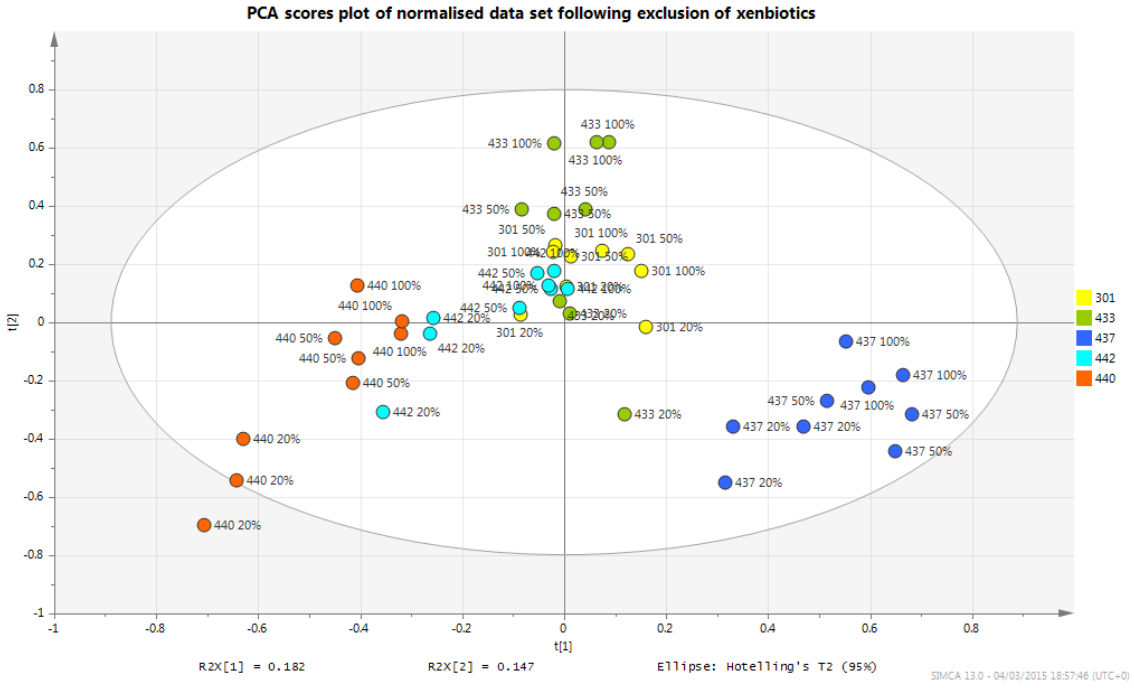


Figure 4.13: Score plot analysis following the exclusion of the 3 identified xenobiotic markers

The effect of xenobiotic exclusion on the PCA scores plot has been minimal compared to the final scores plot from the normalisation section (Figure 4.9). This suggests that endogenous metabolite variation is driving the discrimination between individuals.

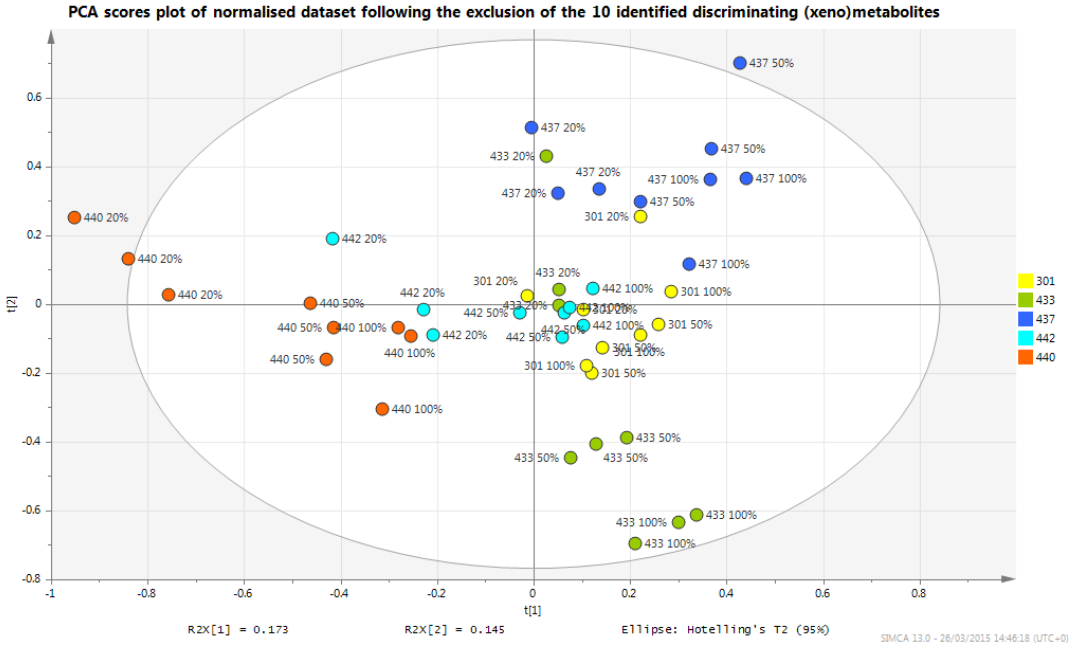


Figure 4.14: Score plot analysis following the exclusion of all 10 identified (xeno)metabolites

Following the exclusion of the identified (xeno)metabolites a significant shift in clustering of scores is observed. However, discrimination between some individuals remains, suggesting other metabolites still drive discrimination between individuals.

4.5 Conclusion

The normalisation method developed in this study offers a potential improvement for the correction of urine concentration. While many previous studies have focussed on post-analysis normalisation, this study compared these methods to a pre-analysis normalisation strategy. The implementation of post-analysis normalisation was unable to correct for serially diluted urine samples. This in part was likely due to metabolites being lost through dilution. The use of a pre-analysis normalisation to osmolality ensures all samples are of equal osmolality prior to LC-MS analysis. This improved individual grouping on a PCA scores plot in addition to increasing the number of common peaks detected. The repeatability of these common peaks was further improved with the addition of a post-analysis MSTS or MSTUS normalisation, which corrected for peak area variation as a result of any LC-MS variability. This normalisation method is also advantageous as no further analyses other than an osmolality reading are required. This proposed method therefore offers a quick and simple normalisation methodology for urine metabolomics. Further work revealed that the impact of xenobiotics on the metabolome of healthy individuals was demonstrated to have a minimal effect on the discrimination of samples between individuals after PCA modelling of the data set and the main driving force of discrimination was found to be numerous endogenous metabolites. However, removal of drug derived xenobiotics maybe more important in a clinical case study,

where the pharmaceutical impact on the metabolome a more important factor to take into consideration during multivariate analyses of the data set.

Chapter 5: A case study investigating the effect of HIV status on metabolomic profiles of urine

Abstract

The world is currently in the midst of a worldwide epidemic of HIV with at least 60 million people having been infected with the virus and 25 million dying as a result. Despite this, HIV infection has been for the most part passed over by metabolomic researchers, despite the opportunity to increase the understanding of the mode of action and effects of retroviral infection. The few studies that have been carried out to investigate the effects of HIV infection on the metabolome have for the most part investigated saliva, blood products and CSF and have typically involved a treatment option. To date, no urine metabolomic study has been performed using LC-MS or other metabolite profiling techniques. Using newly developed sample preparation and LC-MS analytical methodologies, this study investigated the metabolomes associated with three HIV statuses; HIV negative, acute HIV and chronic HIV infections. These new techniques however were unable to detect any discriminating metabolites associated with HIV status or viral load. This result was potentially a result of the effect of HIV on the patients being minimal, as none of the patients reached threshold levels of CD4 count (a marker of T-helper lymphocyte cell numbers) to require antiretroviral therapy (ART) at the time of the study. This would suggest that urine metabolomics for early detection or qualitative analysis of early HIV infection is unsuitable and that other matrices such as blood products maybe a better option despite the inherent risk of infection from HIV retrovirus present in the plasma or serum samples.

5.1 Introduction

The Human Immunodeficiency Virus (HIV) is a current worldwide epidemic with an estimated 60 million people having been infected and 25 million dying as a result (Sharp and Hahn, 2011). The vast majority of cases are due to HIV-1 infection, which is the most widespread and contagious of the two HIV viruses (HIV-1 and HIV-2). Both are zoonotic viruses related to simian immunodeficiency viruses (SIV) which crossed the species border from various primate species in sub-Saharan Africa (Hahn et al., 2000, Sharp and Hahn, 2011, Faria et al., 2014). The HIV virus decimates the immune system of its hosts via its infection of primarily CD4 positive (CD4+) cells of the immune system (Levy, 1993, Pantaleo et al., 1993a). This leads to opportunistic infections and malignancies such as Kaposi's sarcoma (Sharp and Hahn, 2011). These symptoms are typically associated with acquired immunodeficiency (AIDS) as a result of HIV infection and occur when CD4+ lymphocyte cell counts are <200 cells per µL of blood (Pantaleo et al., 1993a, Pantaleo et al., 1993b).

HIV pathogenesis has distinct phases, and upon initial infection patients often experience flu like symptoms which is a result of a dramatic drop in immune function due to loss of CD4+ lymphocytes and peak HIV viral load. This occurs while the patient is undergoing HIV seroconversion and is often referred to as acute HIV infection. After several weeks the host immune system recovers, although never to its previous functionality. However, despite this the immune system is still capable of limiting the replication and spread of the HIV virus and the patient enters a period of clinical latency. During this period, which can last for several years, patients are typically asymptomatic and may not be aware of their HIV status. Once the virus overcomes the host immune system and the CD4+ counts drop to <200 cells per µL of blood, then aggressive treatment with antiretroviral pharmaceuticals ensues. Eventually advanced clinical HIV infection sets in which is characterised by frequent AIDS related opportunistic infections and cancers, and has a very high level of mortality (Pantaleo et al., 1993a, Pantaleo et al., 1993b, Fauci, 1996).

Despite the worldwide spread of HIV and the number of individuals living with the disease, very few metabolomic studies of HIV infection have been carried out. Those that have investigated HIV infection have analysed the plasma, serum, saliva, cerebrospinal fluid (CSF), cell lines and bronchoalveolar lavage metabolomes (Cassol et al., 2014, Philippeos et al., 2009, Hollenbaugh et al., 2011, Wikoff et al., 2008, Ghannoum et al., 2013, Sitole et al., 2013, Williams et al., 2012, Cribbs et al., 2014). To date, the majority of these studies have been on patients taking some form of ART thus making it difficult to differentiate between the metabolomic effects of HIV status and ART intervention. However, one study that used GC-MS analyses to investigate the serum metabolomes of ART naïve HIV positive and negative patients, revealed that a number of disrupted metabolomic pathways were observed. The metabolite profiles of HIV positive patients revealed dysfunction of the mitochondrial system with disruption of concentrations of a number of Krebs cycle organic acids indicating that the respiratory chain in the mitochondria is disturbed. In addition, markers of oxidative stress and disrupted lipid profiles were also elevated in HIV positive patients and this may be a result of an inflammatory response to infection (Williams et al., 2012). An investigation of the effects of HIV infection on the CSF metabolome by Cassol et al., (2014) detected markers of glial activation and disruption to neurotransmitters, ketone bodies, carnitines and amino acids regardless of cART (combined ART medication) status when compared to cART naïve HIV negative patients. These markers were strongly correlated to plasma protein indicators of inflammation, including interleukins and interferons. Similar findings were also found in a metabolomic analysis CSF from macaques infected with the primate analogue of HIV, i.e. SIV.

In addition, a number of carnitines were also found to be disrupted in the samples from SIV infected macaques suggesting lipid metabolism had been disrupted possibly as a result of mitochondrial dysfunction (Wikoff et al., 2008). A targeted metabolomics approach using LC-MSMS investigated the metabolic consequences of HIV infected CD4+ T-cells and a macrophage cell line that is a model for the effect of HIV infection. Interestingly both cells types reacted differently, with glycolysis being increased in the HIV positive CD4+ T-cells and reduced in the macrophages. In addition, the T-cells increased the production of ribose via an oxidative pentose phosphate pathway (Hollenbaugh et al., 2011).

Perhaps surprisingly, to date there have been no urine metabolomic studies of HIV infection. Urine would offer a non-invasive method for metabolomics and has the additional benefit of a minimal risk of HIV infection for the researchers. The aforementioned studies suggest that there are metabolomic consequences of HIV infection that can be detected using a metabolomics approach. The use of improved sample preparation and more sensitive nLC-nESI-MS may aid the discovery of metabolite variation in urine samples. This may include small molecule markers of inflammation such as eicosanoids to compliment the current protein markers of inflammation in plasma (Kamat et al., 2012, Malherbe et al., 2014). Furthermore, no metabolomic study has yet investigated the differences between acute and chronic HIV infection. As such, it is not known how the metabolic response to HIV changes over the course of HIV infection.

5.1.1 Study aims

This study aims to characterise any changes in the urinary metabolome as a consequence of HIV infection. Urine samples from cART naïve HIV negative patients will be analysed in addition to those from cART naïve patients with acute and chronic HIV infection. Using newly developed SPE sample preparation techniques and nUHPLC-nESI-TOFMS methodologies, low as well as high abundance metabolites will be profiled and analysed for metabolomic disruption as a result of HIV infection (Chetwynd et al., 2014, Chetwynd et al., 2015).

5.2 Materials and methods

5.2.1 Materials and chemicals

HPLC grade solvents were purchased from Rathburn Chemicals Ltd (Walkerburn, Scotland, UK) and UHPLC grade solvents from Fisher Scientific UK (Loughborough, UK). Strata X-AW and X-C 60 mg/3ml solid phase extraction (SPE) cartridges were purchased from Phenomenex (Macclesfield, U.K). Deuterated compounds were used as internal standards (IS); 17β-estradiol 2,4,16,16-d4 sodium 3-sulfate (E2-d4-S, >99% D atom), carbamazepine (ring-d10), venlafaxine

(N,N-dimethyl-d6), and diclofenac (phenyl-d4) were purchased from Cambridge Isotope Laboratories Inc. (MA, USA). Progesterone-2,2,4,6,6,17R,21,21,21-d9 (P-d9, 98% D atom) was purchased from CDN isotopes (Quebec, Canada) and prostaglandin E2-d4 (9-oxo-11α,15S-dihydroxy-prosta-5Z,13E-dien-1-oic-3,3,4,4-d4 acid) were purchased from Cayman Chemical Company (MI, USA). All other standards and reagent chemicals were purchased from Sigma-Aldrich Company Ltd., Dorset, U.K.

5.2.2 Sample collection

Urine samples were collected by Brighton and Hove University Hospital (BSUH) from 451 male patients from a HIV clinic. All patients gave written informed consent and ethical approval was given by the NHS Research Ethics Committee (NREC 09/H1107/101). All samples were immediately frozen at -80 °C prior to sample preparation. For this study 54 urine samples were available for analysis from the available cART naïve patients and these comprised 13 HIV negative, 15 acute HIV and 26 chronic HIV samples. Samples were classified as from acute and chronically infected patients as explained below in section 5.3.1. All samples were from patients not receiving medication for HIV, all of whom tested negative for hepatitis B and C. Patients were matched for ages and ethnicity, and Table 1 contains details of mean CD4 and viral counts for the patient groups in addition to several other urine chemistries.

5.2.3 Sample preparation

Urine sample osmolality was measured at BSUH pathology department using a Model 3320 osmometer (Advanced Instruments Inc.) Samples were then stored in 10% methanol and transported to the University of Sussex for metabolomic analysis. Urine samples were defrosted, vortex mixed and centrifuged at 13000 rpm for 10 minutes. Samples were adjusted to pH 2 using formic acid and then spiked with 0.5 ng/µL IS. Samples were extracted using a Strata X-AW stacked on top of a Strata X-C solid phase extraction cartridge which was primed with 1 mL methanol, washed with 1 mL water, and then loaded with 0.5 mL acidified sample before a further 1 mL water wash. Samples were eluted with 1 mL 5% ammonium hydroxide in methanol and a further 1 mL ethyl acetate. Eluents were dried to dryness before reconstitution in 90:10 H2O: MeOH, and the volume for reconstitution were adjusted to ensure all samples had the same calculated osmolality concentration of 1360 mOsm. This concentration was chosen based upon 2 x the mean osmolality of all 54 samples measured prior to SPE. Extracted samples were stored at -80 °C prior to analysis.

5.2.4 nUHPLC-nESI-TOFMS

Aliquots of 0.5 µL were injected onto a Waters nanoAcquity UHPLC and separated using a Waters nanoAcquity HSS-T3 (100mm x 100 µm x 2.8 µm, 100 Å) column. Chromatographic separation was carried out at 700 nL/min using UHPLC grade water and acetonitrile as mobile phases A and B respectively, both modified with 0.01% formic acid. A gradient elution was used: 0mins 10% B, 4mins 30% B, 18mins 50% B, 30mins 100% B, 100% B maintained for 10 minutes then equilibrated in initial conditions for a further 15 mins. Metabolites were detected in positive and negative nESI mode using a Waters Xevo G2 TOFMS tuned to a mass resolution of 20,000 and equipped with a nano ESI source with a homemade fused silica emitter as described in chapter 2 (Chetwynd et al., 2014, Chetwynd et al., 2015).

5.2.5 Quality control

Prior to the analysis of samples in either ionization mode, 5 QC injections were run to condition the ESI source. To determine the repeatability and reliability of the analysis, QC samples were injected every 10 samples. These were then assessed for metabolome peak areas repeatability by calculating the %CV for all peaks present in 80% of QC samples. High quality metabolomic data is determined when 70% of peaks having a mean peak areas of <30% (Want et al., 2010).

5.2.6 Data and statistical analysis

Chromatograms were peak picked and deconvoluted using Waters MarkerLynx software and a MSTS normalisation applied to correct for MS variation. All data were exported for multivariate analysis in Umetrics Simca 13.0 where it was log transformed and Pareto scaled. Principal component scores plots were constructed for both ionization modes, with any discriminating markers being determined using S-plot analysis of the OPLS-DA models. Statistical analysis of patient blood and urine biochemistry was carried out using GraphPad Prism version 6.00 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com. Data from patient groups were tested for normality using the D'Agostino-Pearson omnibus K2 test. Metabolites that were normally distributed were then tested for significance at $P<0.01$ using a one-way ANOVA with a Holm-Sidak multiple comparison test to determine p-values. Non-normally distributed data were tested using the Kruskal-Wallis ANOVA with a Dunn's multiple comparison test to determine p-values.

5.3 Results and discussion

5.3.1 Blood and urine biochemistry

Samples from HIV positive patients were grouped as acute (seroconverters) or chronic infection at the time of donation, based on the Recent Infection Testing Algorithm (RITA) which was carried out at Brighton and Sussex University Hospital. Here the concentrations of various HIV antibodies are used to predict the time elapsed since infection and HIV status. The antibodies used in the determination of RITA scores include immunoglobulin M (IgM) and IgG. IgM is the first antibody to react to a foreign antigen and typically peaks within 2 weeks of HIV infection after which levels drop to initial concentrations, after which IgG levels begin to increase and remain increased during chronic HIV (Murphy and Parry, 2008). RITA uses set cut-off points for these antibody levels to determine HIV acute/chronic status. However, due to inter-individual variation, RITA only offers a rough guide as to when an individual became infected. As such recent infection/acute HIV can be defined for up to 6 months following the initial infection (Murphy and Parry, 2008, Sane et al., 2014).

All samples were from males who were matched for age, urine osmolality and creatinine, urine protein creatinine ratio (UPCR) and cART status. The data sets for each parameter were tested for normality using D'Agostino-Pearson omnibus K2 test and then for significance ($p<0.05$) using one-way ANOVA or for non-normally distributed data a Kruskal-Wallis test. None of the measured parameters varied significantly ($p>0.05$) between the sample groups (Table 5.1). The UPCR is a marker of kidney function with values of >30 indicating impaired kidney function. All of the patients in the sample groups had a UPCR <30. Urine creatinine is used widely as a marker of urine concentration as discussed in Chapter 4, and elevated levels may also be used as an indicator of kidney malfunction. No significant differences were observed between sample groups for either UPCR or urine creatinine indicating there were no changes in kidney function as a result of HIV infection.

The CD4 count and viral load were tested for significance using a two-tailed t-test and showed no significant difference in concentrations between acute and chronic patient groups. All but one patient in the acute HIV group have CD4+ cell counts exceeding 200 $x10^6$ cells/mm^3 and this meant that the HIV infection has not yet reached a point where antiretrovirals are required to suppress the virus. However, CD4 levels were not determined for the HIV negative patient group, but generally in the male population concentrations are 840±285 cells $x10^6/mm^3$ (Bofill et al., 1992). This indicated that CD4+ levels in the cART experienced groups

maybe suppressed compared to a healthy male population, although not to a level requiring antiretroviral therapy.

The viral load was not measured the HIV negative group as they are not expected to have copies of the HIV due to their negative HIV test result, however it was expected to be below 40 RNA copies/mL plasma which is the LOD of the detection method. This would indicate that there was indeed a clear significant increase in viral loads between the HIV negative and positive patient groups. However, it might have been expected that the viral load would have been significantly greater in patients with acute HIV infections as this stage of HIV infection is characterised by peak viral loads and is also referred to as the seroconverion phase (Pantaleo et al., 1993a, Pantaleo et al., 1993b, Fauci, 1996).

Table 5.1: Blood and urine biochemistry of samples from control, acute and chronic HIV patient groups

	HIV -ve	Acute HIV	Chronic HIV
n	13	15	26
Age	34±8	34±8	37±10
CD4 count (cellsx10^6/mm^3)	nd	546±201	507±175
Viral load (RNA copies/mL plasma)	nd	782676±2287956	26711±50205
Osmolality (mOsm)	735±241	659±283	644±281
Creatinine (µmol/L)	14.4±8.9	11±4.5	13±7.7
UPCR	6.7±1.2	8±2.9	10±4.6
cART status	naïve	naïve	naïve

Values are mean ± standard deviation (SD). nd = not determined

5.3.2 Repeatability of nUHPLC-nESI-TOFMS analyses

The peak area repeatability of 5 QC samples was assessed using the method proposed by Want et al., (2010) by calculating the number of peaks present in 80% of the QCs with a %CV of less than 30%. In this study between 70% in negative nESI and 78% in positive nESI of peaks returned a CV of less than 30% thus falling within the acceptable criteria for metabolomics analysis (Want et al., 2010). These values indicate that this metabolomic study is repeatable and thus returned reliable results for further analysis of the data.

5.3.3 Metabolomic analysis of HIV status

The base peak intensity (BPI) chromatograms of a QC sample analysed in positive and negative nESI modes are shown in Figures 5.1 and 5.2 respectively, and several of the most abundant peaks are labelled with the putative identity of the metabolite. Some of these highlighted

peaks cover an important range of metabolite classes such as androgen and mineralocorticoid conjugates (in negative nESI mode) and carnitines (in positive nESI) which play a role in the transport and β-oxidation of fatty acids in the mitochondria. The detection of tryptophan shown in the BPI of Figure 5.1 is of interest due to the previous method development work detailed in Chapter 3 which revealed that analysis by the nano scale platform was crucial for its detection in urine extracts. In addition, examples of selected ion chromatograms of a number of metabolites are also shown that further highlight the sensitivity of this analysis (Figures 5.3 and 5.4). Both testosterone and androstenedione were detected in the urine extracts (Figure 5.3) and these metabolites were previously only detected in extracts after analysis by the nanoflow nanospray platform rather than the ESI platform (Chapter 3). Similarly the detection of the stress response steroid cortisone, which requires nanoflow-nanospray analysis for detection in urine (Chapter 3), was detected alongside cortisol in urine extracts (Figure 5.4) The ability to detect these metabolites is a strong indication that this analysis was highly sensitive and benefited from the improved sample preparation and analytical methodologies developed in Chapters 2 and 3.

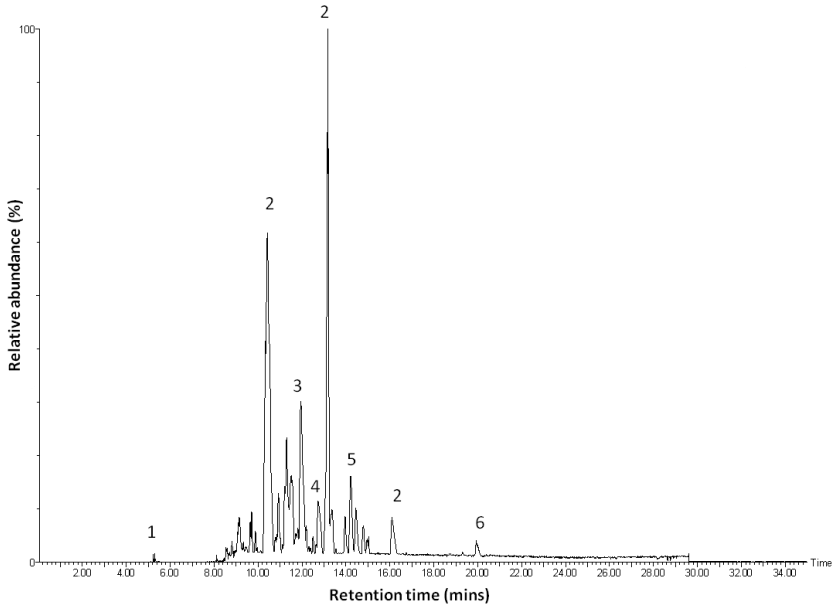


Figure 5.1: Positive nESI BPI chromatogram of the LC-MS analysis of a QC sample of urine extracts

Selected peaks are labelled with the putative metabolite identities; 1) unretained polar metabolites including creatinine and tryptophan, 2) unidentified carnitines, 3) octanoylcarnitine, 4) dimethylheptanoyl carnitine, 5) indolepropionic acid, 6) C17 sphinganine. All carnitine peaks were confirmed from the presence of m/z 85.027 and m/z 60.083 in MS^e data files which are fragment ions common to all carnitine metabolites.

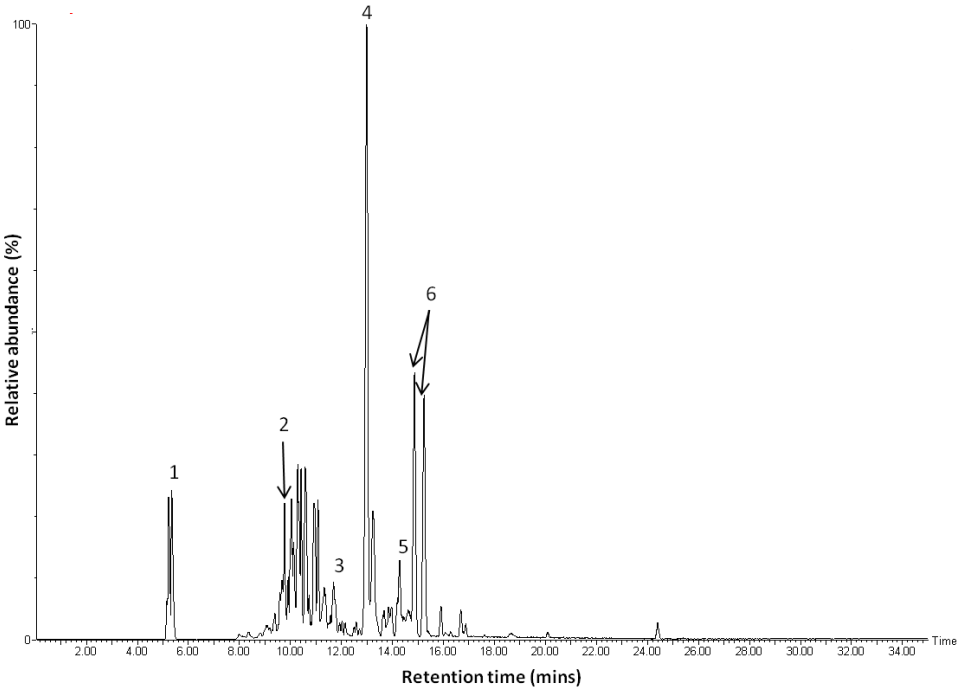


Figure 5.2: Negative nESI BPI chromatogram of the LC-MS analysis of a QC sample of urine extracts

Selected peaks are labelled with putative metabolite identities; 1) unretained polar metabolites including indoxylsulfuric acid, 2) tetrahydroaldosterone glucuronide, 3) testosterone glucuronide, 4) unidentified metabolite 5) androsterone and dihydrotestosterone sulfates, 6) androsterone and dihydrotestosterone glucuronide (unknown elution order).

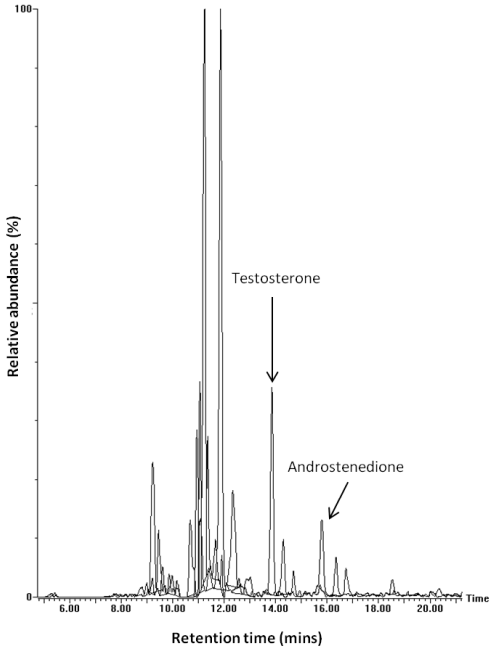


Figure 5.3: Overlaid selected ion chromatograms of unconjugated testosterone and androstenedione detected in urine extracts

The detection of unconjugated testosterone and androstenedione has previously been shown to require nanoflow-nanospray analysis (Chapter 3). Here peaks were detected at *m/z* M+H 289.2168 for testosterone and M+H 287.2011 for androstenedione both within a 5 ppm mass window. The two large peaks eluting between 11 and 12 mins are likely a result of fragmentation of glucuronide conjugates of similar androgen structures. Both metabolite identities were confirmed by retention time comparison to genuine standards and fragmentation patterns (testosterone 271.2054, 253.1932 and 97.0634).

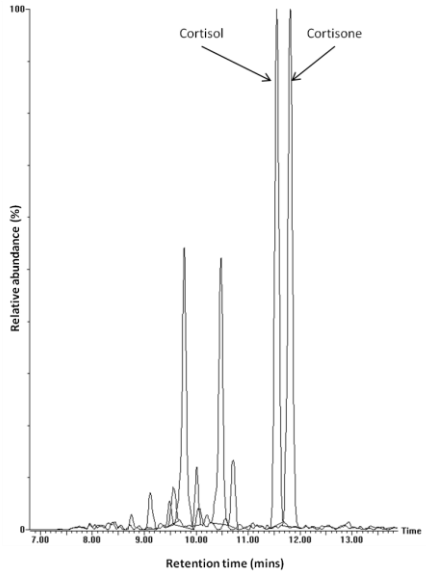


Figure 5.4: Overlaid selected ion chromatograms of unconjugated cortisol and cortisone detected in urine extracts

Peaks were detected at *m/z* M+H 363.2171 for cortisol and M+H 361.2015 for cortisone both within a 5 ppm mass window. The two large peaks eluting between 9.5 and 10.5 mins are likely a result of fragmentation of conjugates of similar steroid structures. Both metabolite identities were confirmed by retention time comparison to genuine standards and fragmentation patterns (cortisol 267.1710) (cortisone 163.1132 and 121.074).

The metabolomic profiles of urine extracts obtained from HIV negative, acute and chronic HIV patients were compared using multivariate statistics. The PCA scores plots generated from the positive and negative ESI data are given in Figures 5.5 and 5.6 respectively. In both ionization modes the models are poor with only between 11.8% and 12.3% of variation explained by the first two components of the scores plot. Neither of the sample groups displays discrimination on the PCA scores plot, and when the data was analysed in a supervised PLS-DA model, the descriptive statistics were poor and require a forced first two components. The observed variation between the samples appeared to be due to inter-individual variations as opposed to HIV status.

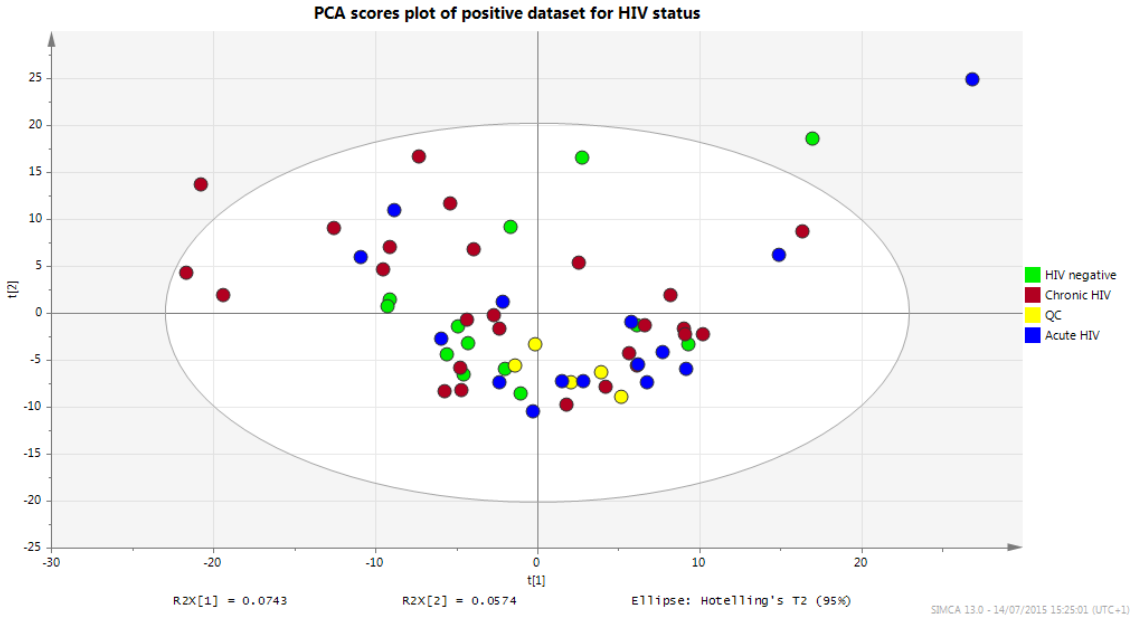


Figure 5.5: PCA scores plot of the effect of HIV status on the urinary metabolome analysed by LC-MS in positive ESI mode

No discrimination was discerned between sample groups suggesting that no significant variation in the urinary metabolome as a result of HIV infection was detected.

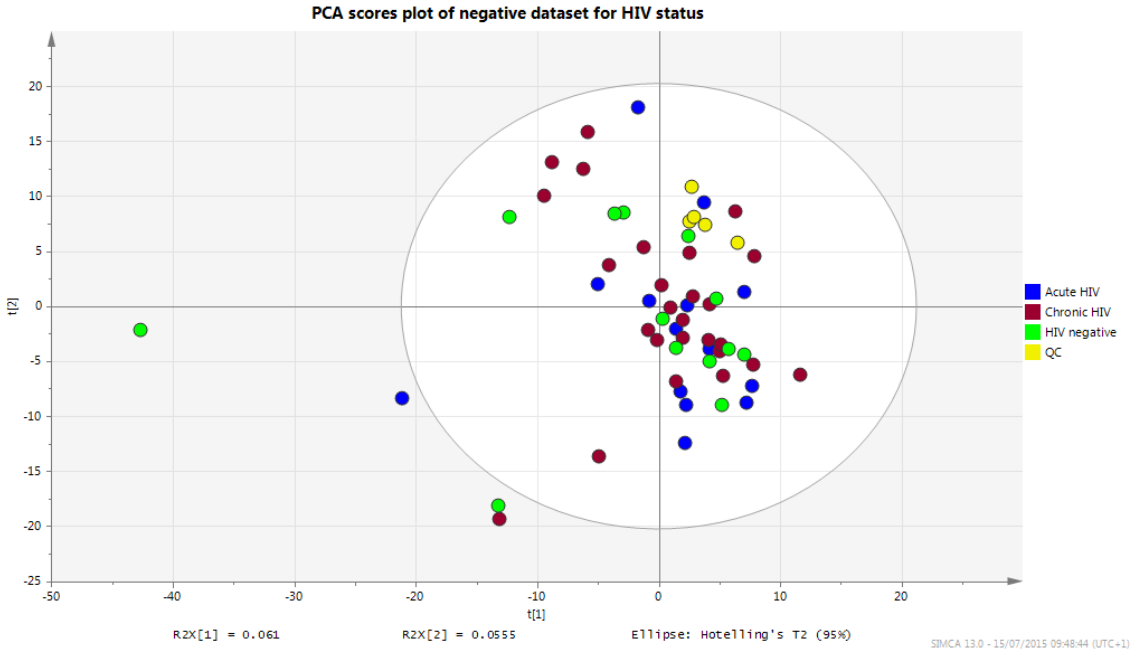


Figure 5.6: PCA scores plot of the effect of HIV status on the urinary metabolome analysed by LC-MS in negative ESI mode

No discrimination was discerned between sample groups suggesting that no significant variation in the urinary metabolome as a result of HIV infection was detected.

An additional consideration is that the viral load within the samples grouped as either acute or chronic stages of HIV infection varied significantly (Table 5.1). To examine the effect of viral load on urinary metabolite profiles, the data sets from the acute and chronic HIV groups were combined and the data remodelled by stratifying the groups by viral load. In addition to the 13 HIV negative samples, 9 samples were defined as high viral load of between 49,000x10^6 and 2x10^6 HIV RNA copies/mL plasma, and 15 samples with a low viral load of between 100 and 5000 HIV RNA copies/mL plasma. The viral load of plasma from HIV negative patients was not determined however it was likely to be less than 40 RNA copies/ mL plasma which was the limit of detection of the method (Wang et al., 2010a, Cobb et al., 2011)Modelling of these sample groups using the negative nESI data set proved to be impossible for either unsupervised PCA or supervised PLS-DA as these models required forcing of the first component. In positive nESI a model was successfully constructed however no discrimination between sample groups was evident. A forced PCA scores plot model for the negative nESI data where two components were fitted to the dataset and an unforced positive nESI PCA scores plot are shown in Figures 5.7 and 5.8 respectively.

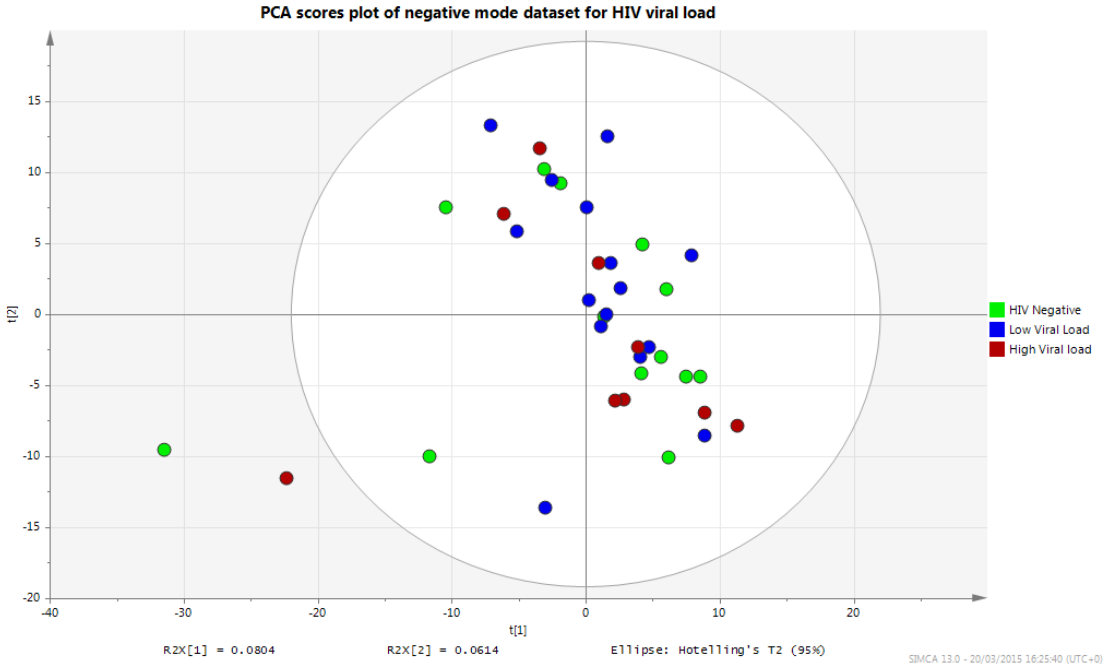


Figure 5.7: PCA scores plot of the effect of HIV viral load on the urinary metabolome analysed by LC-MS in negative ESI mode

No discrimination between sample groups was discerned suggesting no significant variation in the urinary metabolome of patients was detected that was associated with their viral load.

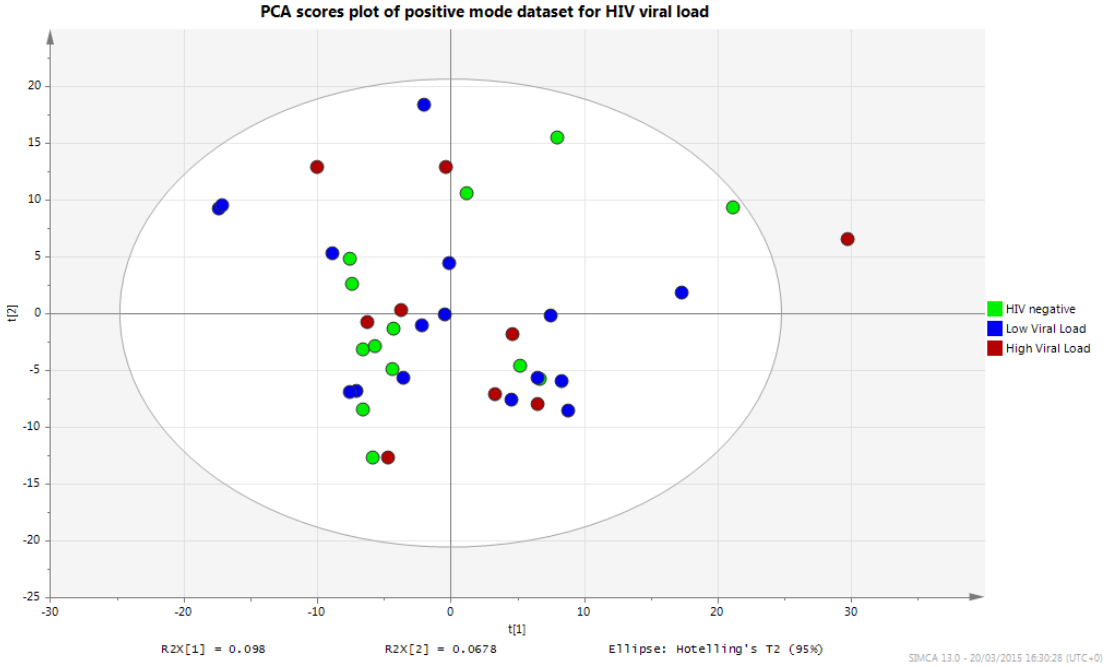


Figure 5.8: PCA scores plot of the effect of HIV viral load on the urinary metabolome analysed by LC-MS in positive ESI mode

No discrimination between sample groups was discerned suggesting no significant variation in the urinary metabolome was detected that was associated with patient viral load.

The lack of discrimination between sample groups, defined either by acute/chronic HIV infection or by viral load, may be due to the early stages of HIV infection. All but one HIV patient (in the acute HIV group) had a CD4+ count greater than 200 cells/mm^3. This meant that no patients in this study required antiretroviral therapy at this point of their HIV progression as the virus had not yet disrupted immune function to the extent of requiring pharmaceutical intervention. The patient with the low CD4+ count may subsequently require cART intervention however at the time of sample collection they were cART naïve. New guidelines that suggest treatment should commence once the CD4+ cell count drops to $350x10^6$ cells/ mm^3, this would increase the number of patients that may go on to be prescribed cART in this study to 6 (2 patients from the acute and 4 chronic group) (Williams et al., 2014). The previous metabolomic studies on samples from cART naïve HIV patients had investigated either blood products, which contain the HIV retrovirus (Williams et al., 2012) or investigated neuroAIDs, an advanced condition associated with HIV in the CSF (Cassol et al., 2014, Hollenbaugh et al., 2011). In these cases the effect of HIV infection on the metabolome is likely to be more profound due to the blood products carrying the virus itself or the more advanced stage of HIV infection in the patients. To the author's knowledge this is the first LC-MS metabolomic analysis of the urinary metabolome of individuals with HIV, and indicates that analysis of this biofluid may not be as informative as blood samples in determining the wider effects of HIV on metabolite profiles.

This study aimed to analyse the profiles of low abundance alongside high abundance metabolites in the urine and which are potentially affected as a result of HIV infection. As HIV is an inflammatory disease (Deeks, 2011, Hunt, 2012), then profiles of small molecule markers of inflammation such as the eicosanoids maybe expected to be disrupted in the plasma and therefore in the urine. Several protein inflammatory markers are known to be increased in patients with HIV such as pro-inflammatory cytokines, interleukins, tumour necrosis factors and C-reactive protein (Decrion et al., 2005, Hunt, 2012). Many studies have reported on disrupted eicosanoid levels during HIV infection. This disruption includes increased synthesis of PGE2 and PGB2 in infected monocytes (Foley et al., 1992, Conant et al., 1998, Ramis et al., 1991). Enhanced CSF concentrations of prostaglandin F2α, prostaglandin D2, leukotriene B4 and thromboxane B2 have been observed as a result of advanced HIV infection (Adamson et al., 1996, Griffin et al., 1994, Froldi et al., 1992). In addition, urinary concentrations of prostaglandin E2 metabolites and 8-iso-prostaglandin F2α were increased in HIV infected women compared to HIV negative (Fitzgerald et al., 2012, Boger et al., 2012). In plasma, F2-isoprostane concentrations been observed to decrease when patients are receiving

antiretroviral therapy (Hulgan et al., 2003). These studies indicate that eicosanoid concentrations in several biofluids and cells are known to be disrupted as a result of HIV infection. However, in this study multivariate analysis of the LC-MS data sets did not reveal any disruption of eicosanoid metabolites including prostaglandins in the urine extracts. The poor detection of eicosanoids was possibly as a result of the detection of non-steroidal anti inflammatory drugs (NSAIDs) in the urine which is discussed below. Despite this, a manual search chromatograms of urine extracts from some of the patients not taking NSAIDs did return potential eicosanoid peaks. These potential eicosanoids included tetranor PGE-M (*m/z* 327.144), 11-dehydrothromboxane B2 and a glucuronide conjugate (*m/z* 367.200 and 543.244 respectively), 20-hydroxyleukotriene B4 and a glucuronide conjugate (*m/z* 351.217 and 527.249 respectively). However, these peaks were very low abundance and it was not possible to confirm the identities through fragmentation patterns. Further targeted analysis would be required to ascertain the identities of these peaks.

During urine collection, only one patient reported consuming an NSAID, however many urine samples from individuals who did not report taking NSAIDS were found to contain NSAIDs. The parent NSAIDs that were detected included paracetamol, ibuprofen and naproxen all of which were confirmed by comparison of retention times and fragmentation patterns with genuine standards (Table 5.2). Exposure of patients to these pharmaceuticals would reduce the chances of detecting any variation in eicosanoid levels as a result of HIV infection. This is due to the action of NSAIDs which inhibit the enzymes cyclooxygenase 1 and 2 (also termed PGH synthase 1 and 2) which are responsible for production of eicosanoids (Hawkey, 1999, Chandrasekharan et al., 2002). Analysis of the chromatograms for all patients revealed that urine extracts from 44 (12 HIV negative, 12 acute HIV and 20 chronic HIV infections) of the 54 patients contained the parent compounds of ibuprofen and an additional patient in the chronic infection group contained naproxen and another contained paracetamol (acute infection group) un-metabolised parent compounds. The fact that most patients did not report taking NSAID suggests that they do not consider common painkillers to be pharmaceuticals. This may be a result of the common misconception that non-prescription over the counter drugs are safer than those requiring a prescription (Wilcox et al., 2005). In addition, patient recall for prescription pharmaceutical has been reported to be a poor indicator of actual pharmaceutical intake (West et al., 1995). This phenomenon is magnified when reported intake of different pharmaceutical classes is investigated, with self-reported prescription painkiller use being the least accurate (Skurtveit et al., 2008). This disparity between actual and reported pharmaceutical intake is also likely to be true of non-prescription pharmaceutical

use. The under reporting of drug use complicates metabolomic and epidemiological studies where additional intake of unreported drugs may influence the outcome of the study. Ideally reported drug use would be accurate so that patients could be stratified by pharmaceutical intake. This would not only improve the quality of the results for the disease state being studied but also provide additional information on the impact of commonly consumed pharmaceuticals and their mixtures on the disease state.

Table 5.2: Details of NSAIDs that were detected in the urine of study groups

NSAID	Exp. Mass	Thr. Mass	Formula	Fragments
Ibuprofen	M-H 205.1225	M-H 205.1229	$C_{13}H_{18}O_2$	161.916
Naproxen	M-H 229.0859	M-H 229.0865	$C_{14}H_{14}O_3$	185.0966, 170.0731, 169.0651
Paracetamol	M+H 152.0716	M+H 152.0712	$C_8H_9NO_2$	150.0991 110.0699

Exp. = experimental, Thr. = theoretical. All metabolite identities were confirmed by comparing fragmentation patterns and retention times to that of pure standards.

A further co-morbidity with men infected with HIV is hypogonadism leading to erectile dysfunction. This has been demonstrated to be due to a result of reduced serum testosterone levels (Rabkin et al., 1997, Crum-Cianflone et al., 2007, Crum et al., 2005). Despite the detection of a number of androgen metabolites including testosterone in the urine extracts (see Figure 5.2 and 5.3) no differences in androgen profiles were detected between subject groups. However, hypogonadism and reduced serum androgen concentrations are associated with low CD4+ counts , advanced HIV infection and age (Wagner et al., 1995). This suggests that these symptoms had not yet presented themselves in these patients due to their relatively healthy CD4+ count. The fact that all these patients are cART naïve regardless of HIV status suggest that the effect of the retrovirus on the body is still in its latent stage or that the effects of the large viral load present in the blood of some patients are not yet reflected in the urinary metabolome.

5.4 Conclusion

This was the first study into the effect of HIV status on the urinary metabolome using LC-MS. Despite the use of new sample preparation and analytical methods designed to detect variability in even low abundance metabolites, no significant differences were detected between sample groups. However, analysis of the urine revealed that most of the patients in the sample groups were taking NSAID medication which may have masked some potential discriminating metabolites which are markers of the inflammatory response. A further explanation is that the patients had not suffered significant metabolic changes in the urine as a result of their HIV status or viral load. This is supported by the fact that none of the patients had begun receiving antiretroviral therapy and that their CD4+ cell count exceeded 200 cells/

mm^3. These results may also indicate that any changes that are present in the body are not reflected in the urine and that metabolomic analysis of blood products such as plasma or serum are more suited to investigating the effects of HIV infection in patients with recently acquired HIV infection. The use of urine metabolomics may be better suited for later stage HIV infection or metabolomic analysis of cART intervention.

Chapter 6: The metabolomic consequences of cART treatment in HIV positive patients

6.1 Abstract

Since the introduction of antiretroviral therapy (ART) the mortality associated with HIV/AIDS has significantly dropped. As a result the number of people living with HIV has risen and these people are on ART for longer, as life expectancy increases with each successive generation of ART drugs. As such, the side effects of these drugs are becoming more significant due to their effect on health and quality of life. Several of these drugs have been associated with kidney damage, liver failure, dyslipidemia and insulin resistance. The use of a urine metabolomics approach using newly developed analytical methodologies offers the potential to study metabolomic consequences of combined anti-retroviral therapy (cART). In this study, the metabolomic effects of two cART regimes were compared to cART naïve patients. This was the first non-targeted metabolomic study of human urine to detail the extensive metabolism of 3 protease inhibitors; atazanavir, darunavir and ritonavir. This may lead to further investigations into links between pharmacogenetics and metabolomics potentially leading to an improved provision of personalised healthcare. Statistically significant variation in endogenous markers associated with bile acid profiles and fatty acid metabolism were also uncovered. These markers offer a potential explanation for the poor lipid profiles associated with the use of protease inhibitors. In addition, a known marker of immune deficiency in urine was found to be significantly lowered in patients receiving treatment suggesting that the cART intervention is efficacious. The use of urine metabolomic profiling allows for the identification of markers of cART intervention allowing the efficacy, compliance and side effects of pharmaceutical intervention to be investigated, in addition to investigating the metabolism of the cART drugs themselves.

6.2 Introduction

Ever since the discovery of the HIV retrovirus in 1983, many pharmaceutical strategies have been taken to treat those afflicted. This has led to the development of >25 different anti-retroviral drugs spanning 8 drug classes (Deeks and Phillips, 2009). This development of anti-retroviral drugs has significantly reduced the mortality and incidence of AIDS related conditions. In addition, a significant drop in opportunistic infections in HIV patients has been observed (Deeks and Phillips, 2009, Levy, 1993). Despite these advances, combined antiretroviral therapy (cART) intervention never fully restores health. However many individuals infected with HIV now die as a result of conditions previously thought to be non-HIV related such as cardiovascular disease, cancer and kidney or liver failure. These conditions

are usually associated with the normal healthy aging process, which has been shown to occur at an advanced rate in HIV patients compared to healthy individuals (Deeks and Phillips, 2009).

The near normal life span of relatively good health comes at a cost which are the side effects of cART medication, and this has become an important issue to contend with due to its effects on patient health and quality of life. The current recommendation for cART therapy is two nucleoside reverse transcriptase inhibitors (NRTI) and a protease inhibitor (PI). This is typically co-prescribed with another protease inhibitor, ritonavir, to boost the bioavailability of the other PI (Williams et al., 2014, Hill et al., 2009, Wood and Flexner, 1998, Zeldin and Petruschke, 2004). All cART drug classes have different associated side effects, some of which are more serious than others. Prior to the introduction of cART, the major cause of nephropathy was as result of the HIV infection itself (Hall et al., 2011a). However, currently the leading cause of nephropathy in HIV patients is a result of tenofovir, a commonly prescribed NRTI. The site of tenofovir toxicity is the mitochondria located on the baso-lateral membrane of the proximal tubule cells of the kidney (Hall et al., 2011a, Kohler et al., 2009). Another drug class, the PIs, are commonly co-prescribed with NRTIs as part of a cART regime. These are widely associated with non-HIV associated complications such as dyslipidemia, abnormal fat distribution, diarrhoea, increased inflammation markers and insulin resistance (Tomaka et al., 2009, Aberg et al., 2012, Wood and Flexner, 1998). Due to the rapid metabolism of PIs by cytochrome P450 3A4, patients take two protease inhibitors. One, ritonavir, is taken as a cytochrome P450 3A4 inhibitor and has been demonstrated to increase the bioavailability of the other PI pharmaceutical by up to 350% (Wood and Flexner, 1998, Hill et al., 2009, Hsu et al., 1997). This inhibition has no effect on other cART classes such as NRTIs which are metabolised by other enzymes (Hsu et al., 1997). Ritonavir itself is associated with all of the aforementioned PI side-effects. In addition to ritonavir, a further PI is prescribed and two of these possibilities are darunavir and atazanavir. These latter two drugs were introduced at least 7 years after ritonavir and have fewer side effects associated with them. Both of these drugs have improved lipid profiles compared to older PIs when tested on cART naïve HIV positive or negative patients (Aberg et al., 2012). Atazanavir and darunavir are typically prescribed when patients suffer from significant side effects such as dyslipidemia when on other older PI medications (Tomaka et al., 2009). However, atazanavir intake is also associated with hyperbilirubinemia, which whilst usually asymptomatic may lead to jaundice and in rare cases Gilbert's syndrome where high levels of bilirubin build up in the blood. This is due to the atazanavir mediated inhibition of uridine 5'-disphospho-glucuronosyltranferase which is used to conjugate bilirubin with glucuronic acid to facilitate its excretion (Rekić et al., 2011, Zhang et al., 2005). However,

these symptoms resolve themselves when atazanavir is replaced with another PI such as darunavir (Aberg et al., 2012, Tomaka et al., 2009).

The use of a non invasive urine metabolomics approach may help to further elucidate mechanisms of action and toxicity associated with cART intervention (Sitole et al., 2013). To date, few studies have investigated the effect of HIV infection or the effect of cART intervention on the metabolome. One study analysed the plasma metabolomes of healthy controls compared to patients with advanced HIV that were receiving antiretroviral therapy incorporating PIs, and some of whom were also co-infected with hepatitis C (HCV) (Cassol et al., 2013). All HIV positive patients had increased concentrations of bile acids, amino acids, nucleotide metabolites and decreased plasma concentrations of sulfated androgens, eicosanoids and poly unsaturated fatty acids (PUFAs). These disrupted metabolite concentrations were detected regardless of the patients HCV status, implying the effect of HCV is masked by the metabolomic effect of HIV and cART. Many of these metabolites such as the bile acids, PUFAs and androgens act as ligands for nuclear receptors and thus impact upon gene expression, and the genes affected by these ligands mostly relate to fatty acid metabolism and inflammation. In addition, increased bile acid concentrations in plasma were found to be proportional to the degree of hepatic fibrosis despite there being no significant difference in bile acid profiles between HCV and non HCV patients. Mitochondrial dysfunctions, as possible side effect of tenofovir medication, were also correlated to the disruption in carnitine metabolism (Cassol et al., 2013). A further study investigated the oral cavity metabolome of patients on cART and compared it to the metabolome of cART naïve HIV positive patients. Here carnitines were again found to discriminate between the two populations in addition to several amino acids, nucleotides and dipeptides. The amino acid and nucleotide disruption was attributed to opportunistic infection of the oral cavity as a result of HIV infection. However, a 0.71 fold increase in glycyltyrosine and a 3.89 fold decrease in leucylisoleucine polypeptide concentrations were unexplained. An increased phenylalanine: tyrosine ratio was found to be a good indicator of an impaired immune system and potentially linked to viral replication as it was suppressed in cART experienced patients who had reduced viral loads. This ratio was suggested as a mechanism to assess cART compliance in HIV positive patients (Ghannoum et al., 2013).

6.2.1 Study aims.

To date, no LC-MS metabolomic study of cART intervention has been undertaken using urine despite urine metabolomics offering an insight into the metabolism of pharmaceuticals and their effect on the metabolome (Kell, 2006).

This study aims to investigate the effect of two cART regimes on the urinary metabolome when compared to cART naïve patients. All cART patients took the same NRTI combination of tenofovir, emtricitabine and PI booster ritonavir. The two cART treatment groups were split by the nature of the second PI which was either darunavir or atazanavir.

6.3 Materials and methods

6.3.1 Materials and chemicals

HPLC grade solvents were purchased from Rathburn Chemicals Ltd (Walkerburn, Scotland, UK) and UHPLC grade solvents from Fisher Scientific UK (Loughborough, UK). Strata X-AW and X-C 60 mg/3ml solid phase extraction (SPE) cartridges were purchased from Phenomenex (Macclesfield, U.K). Deuterated compounds were used as internal standards (IS); 17β-estradiol 2,4,16,16-d4 sodium 3-sulfate (E2-d4-S, >99% D atom), carbamazepine (ring-d10), venlafaxine (N,N-dimethyl-d6), and diclofenac (phenyl-d4) were purchased from Cambridge Isotope Laboratories Inc. (MA, USA). Progesterone-2,2,4,6,6,17R,21,21,21-d9 (P-d9, 98% D atom) was purchased from CDN isotopes (Quebec, Canada) and prostaglandin E2-d4 (9-oxo-11α,15S-dihydroxy-prosta-5Z,13E-dien-1-oic-3,3,4,4-d4 acid) were purchased from Cayman Chemical Company (MI, USA). All other standards and reagent chemicals were purchased from Sigma-Aldrich Company Ltd., Dorset, U.K.

6.3.2 Sample collection

Urine samples were collected by Brighton and Sussex University Hospital from 451 male patients in a HIV clinic. All patients gave written informed consent and ethical approval was given by the NHS Research Ethics Committee (NREC 09/H1107/101). Samples were immediately frozen at -80 °C after collection prior to sample preparation. For this study, 89 samples were analysed, 30 HIV positive patients with the protease inhibitor darunavir, 20 HIV positive patients prescribed atazanavir, 26 cART naïve HIV positive patients and 13 cART naïve HIV negative patients. All cART patients were also taking the same NRTI combination of tenofovir, emtricitabine and PI booster ritonavir. The remaining 362 patients were taking a wide range of ARTs or also suffered from conditions such as hepatitis, a known cause of liver damage and so were not included in this study. The characteristics of the pharmaceuticals used in cART are given in Table 6.1 and their structures in Figure 6.1. Ritonavir is taken as a PI booster, as the bioavailability of other PIs such as atazanavir (measured as its pharmacokinetics in blood plasma) increased by 3 fold when co-prescribed with ritonavir (Le Tiec et al., 2005). Likewise the bioavailability of darunavir is increased from 37% to 82% when boosted with ritonavir (Rittweger, 2007).

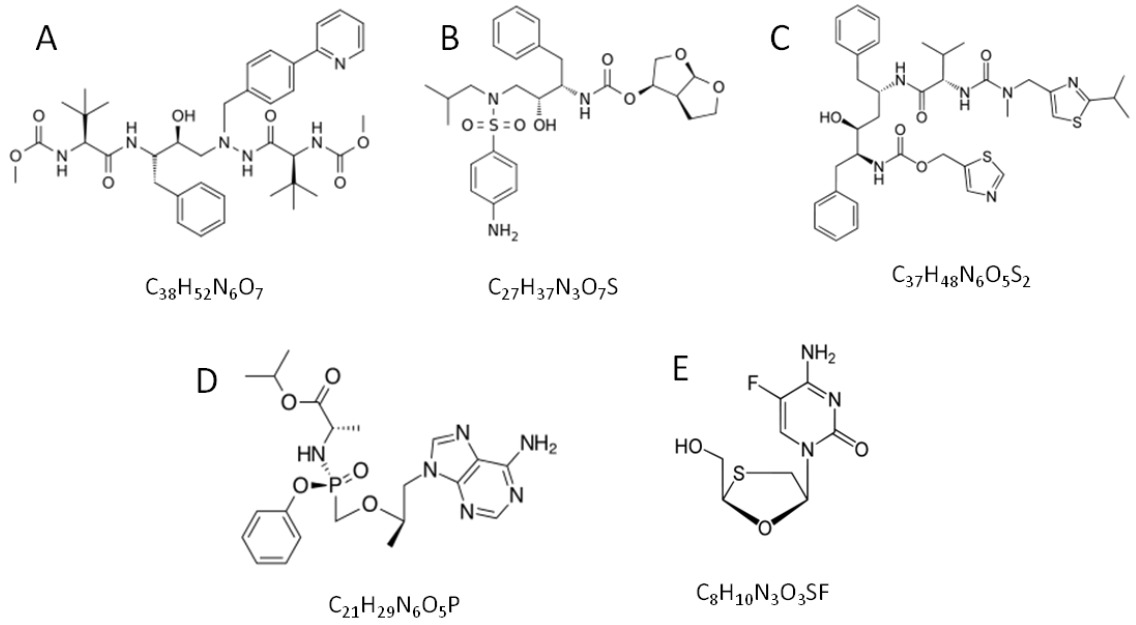


Figure 6.1: Structure and formula for each cART in the study

PIs A) Atazanavir, B) Darunavir, C) Ritonavir, and NRTIs D) Tenofovir, E) Emtricitabine

Table 6.1: Characteristics of NRTI and PI pharmaceuticals used in the study

Antiretroviral	Drug class	Year approved	Log$K_{o/w}$	Water solubility (mg/mL)	Dose (mg)	Half life (hrs)	% urine excretion	Take with food?
Tenofovir	NRTI	2001	-1.71	13.4	1x 300	17	80	Yes
Emtricitabine	NRTI	2003	-0.41	112	1x 200	10	86	N/A
Ritonavir	PI	1996	5.28	0.1	1x 100*	3-5	11	Yes
Atazanavir	PI	2003	5.2	4-5	1x 300	6.5	13	Yes
Darunavir	PI	2006	3.94	0.2	1x 600	15	14	Yes

*When taken with darunavir, a 2 x 100mg or 1 x 200 mg dose is required for PI boosting. Emtricitabine can be taken with or without food with no adverse effect in efficacy or side effects.

6.3.3 Sample preparation

Urine sample osmolality was measured in the BSUH pathology department using a Model 3320 osmometer (Advanced Instruments Inc., Massachusetts, U.S.A). Samples were then stored in 10% methanol and transported to the University of Sussex for metabolomic analysis. Urine samples were defrosted, vortex mixed and centrifuged at 13000 rpm for 10 minutes. Samples were adjusted to pH 2 using formic acid and then spiked with 0.5 ng/µL of IS. Samples were extracted using a Strata X-AW stacked on top of a Strata X-C solid phase extraction cartridge which was primed with 1 mL methanol, washed with 1 mL water, and then loaded with 0.5 mL acidified sample before a further 1 mL wash of water. Samples were eluted with 1 mL 5% ammonium hydroxide in methanol and a further 1 mL ethyl acetate. Eluents were dried to dryness before reconstitution in 90:10 H2O: MeOH. As described in chapter 5 the volume for

reconstitution of each sample was calculated to give an osmolality of 1360 mOsm calculated from pre-extraction readings. Extracted samples were stored at -80 °C prior to analysis. All solvents used were HPLC grade with >99% purity. A QC sample was prepared by combining 2 µL aliquots from all reconstituted urine samples.

6.3.4 nUHPLC-nESI-TOFMS

Aliquots of 0.5 µL were injected onto a Waters nanoAcquity UHPLC and separated using a Waters nanoAcquity HSS-T3 (100mm x 100 µm x 2.8 µm, 100 Å) column. Chromatographic separation was carried out at 700 nL/min using UHPLC grade water and acetonitrile as mobile phases A and B respectively, both modified with 0.01% formic acid. A gradient elution was used: 0mins 10% B, 4mins 30% B, 18mins 50% B, 30mins 100% B, 100% B maintained for 10 minutes then equilibrated in initial conditions for a further 15 mins. Metabolites were detected in positive ESI using a Waters Xevo G2 TOFMS tuned to a mass resolution of 20,000 and equipped with a nano ESI source with a homemade fused silica emitter as described in recent studies (Chetwynd et al., 2014, Chetwynd et al., 2015).

6.3.5 Quality control

Prior to the analysis of samples in either ionization mode, 5 QC injections were run to condition the ESI source. To determine the repeatability and reliability of the analysis, QC samples were injected every 10 samples. These were then assessed for metabolome peak areas reproducibility by calculating the %CV for all peaks present in 80% of QC samples. High quality metabolomic data is determined when the CV of the mean peak area for more than 70% of peaks is <30% (Want et al., 2010).

6.3.6 Data and statistical analysis

The MS datasets were deisotoped, mass spectral peaks deconvoluted, aligned, and the datasets binned and normalized to the total spectral area for each sample using Waters MarkerLynx software. The decision to use MSTS was taken due to only a small proportion of peaks being common to all chromatograms making the use of MSTUS normalisation undesirable. The datasets, comprising RT x m/z bins, were exported to Simca v13.0 software (Umetrics Ltd, Crewe, UK) for multivariate analyses. All data were log transformed and Pareto scaled prior to principal components analysis (PCA) to identify the effect of cART intervention. Further modelling using orthogonal partial least-squares discriminant analysis (OPLS-DA) was used to investigate metabolite differences between any two treatment groups. Discriminatory metabolites (loading variables) between treatment groups were detected using an S-plot of the OPLS-DA model which is a plot of reliability (correlation) of the loading variables versus their covariance (contribution to the model) (Wiklund et al., 2008).

Further statistical analysis was carried out using GraphPad Prism version 6.00 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com. The normalised mass intensities for discriminatory metabolites highlighted by S-plot analysis were tested for normal distribution of the data set and were then tested for significance at $P<0.01$ using a one-way ANOVA with a Holm-Sidak multiple comparison test to determine p-values. Non-normally distributed data were tested using the Kruskal-Wallis ANOVA with a Dunn's multiple comparison test to determine p-values. Statistically significant metabolites were tested for their predictability using Receiver-Operator Characteristic (ROC) to plot the false positive rate against the true positive rate and the area under the curve was calculated to give the probability of a positive result being a true positive.

Metabolite identities were determined from their accurate mass, isotopic fit, and comparison of fragmentation data with authentic standards or with Metlin (Tautenhahn et al., 2012a), Human Metabolome Database (Wishart et al., 2007), Human Urine Metabolome Database (Bouatra et al., 2013) and MycompoundID (Li et al., 2013a) databases.

6.4 Results and discussion

6.4.1 Patient data

All patients in the cART groups took the two NRTIs tenofovir and emtricitabine, and the ritonavir protease inhibitor, as well as either one of the two additional PIs which were either atazanavir or darunavir. These patients all exhibited good viremic control with their viral loads below the limit of detection of 40 RNA copies/ml plasma (Table 6.2). HIV negative patients did not have CD4 counts or viral loads determined as they are not expected to have the HIV virus present due to their negative HIV test, thus their CD4 counts are expected to be unaltered and around the average for a healthy male of 840±285 cells $x10^6/mm^3$ (Bofill et al., 1992).

Each parameter was tested for normality using the D'Agostino -Pearson omnibus K2 test, non-normally distributed datasets were then tested for significance ($p<0.05$) using one-way ANOVA and Dunns multi comparison test. The normally distributed dataset for creatinine and osmolality were tested for significance using a parametric one-way ANOVA with Holm-Sidak multiple comparison test.

Patients in the cART experience groups were significantly older than both cART naïve groups, this is potentially due to these patients having been infected with HIV for a longer period of time and as such their HIV condition now requires cART intervention. The viral loads of the cART naïve patients are significantly higher than all other groups due to the fact they are untreated. In addition the urine protein creatine ratio (UPCR) is significantly increased in cART

experienced patients relative to the HIV negative patients; further to this the patients in the atazanavir group have a significantly greater UPCR relative to the patients in the cART naïve HIV positive group. The UPCR measure is an indication of kidney function and a value of >30 indicates that kidney damage may have occurred. While the values in this study are not indicative of poor kidney function, it may reflect the HIV or drug associated nephropathy was beginning to manifest sub-clinically in these HIV positive patient groups. The urinary osmolality and creatinine levels were not significantly different between any of the sample groups. Atazanavir is well documented to cause subclinical hyperbilirubinemia (Rekić et al., 2011), in this study plasma concentrations of bilirubin in the atazanavir group were significantly ($p<0.05$) greater than the other 3 sample groups.

Table 6.2: Blood and urine biochemistry of samples from cART naïve HIV positive and negative patient groups in addition to cART experiences patient groups

	HIV -ve	HIV +ve	Atazanavir in cART mix	Darunavir in cART mix.
n	13	26	20	30
Age	34±8	37±10	46±6.7*	48±7.5*
CD4 count (cells x10^6/mm^3)	nd	507±175	608±267	529±187
Viral load (RNA copies/mL plasma)	nd	26711±50205^{+}	40±0	40±0
Osmolality (mOsm)	735±241	644±281	700±178	577±259
Creatinine (µmol/L)	14.4±8.9	13±7.7	13±4.7	12±8
UPCR	6.7±1.2	10±4.6	14.5±8.33$^{-\#}$	14±7^{-}
Plasma bilirubin (µmol/L)	12.4±3.1	7.9±2.7	32.5±15.4$^{\sim}$	7.4±2.9
cART status	naïve	naïve	cART positive	cART positive

Values are mean ± standard deviations. * Significantly older than cART naïve groups, $^{+}$significantly greater viral load compared the rest of the patient population. $^{-}$ UPCR levels are significantly greater for both cART experienced groups compared to HIV negative patients. $^{\#}$The atazanavir cART group has a significantly greater UPCR relative to the cART naïve HIV positive group. $^{\sim}$Plasma bilirubin concentrations in the atazanavir group were significantly greater than all other groups. In all cases significance is defined by a p-value <0.05. nd= these values were not determined in HIV negative patients

6.4.2 Repeatability of the nUHPLC-nESI-TOFMS metabolomic analyses of the QC samples.

Using the criteria in previous chapters, the repeatability of metabolome peak area was calculated by determining the percentage of peaks common to 80% of the QC samples and which had a mean peak area CV <30% (Want et al., 2010). In this study between 68% (in negative nESI mode) and 75% (in positive nESI mode) of these peaks have a CV <30% which is acceptable with the repeatability of the negative ESI data set being slightly lower than the positive ESI data set.

6.4.3 Metabolomic analysis of the effect of cART treatment.

The urine metabolomic profiles of individuals taking the two cART regimes were compared to HIV positive and negative cART naïve patients. The scores plots for these data in positive and negative ESI are shown in Figures 6.2 and 6.3 respectively, both display tight grouping of the QC samples further supporting the high quality nature of the metabolomic analysis. As with the previous study in Chapter 5 there was no discrimination detected between the metabolomics profiles of urine from cART naïve HIV positive and negative patients. However, patients on cART regimes containing either atazanavir or darunavir clustered separately and away from the cART naïve patients. In positive ESI, the two treatment groups are separated from the cART naïve groups on the first component and from each other on the second component. However, in negative ESI discrimination is only observed on the first component, which suggests that the two nESI modes are detecting different metabolomics profiles associated with the two types of cART intervention.

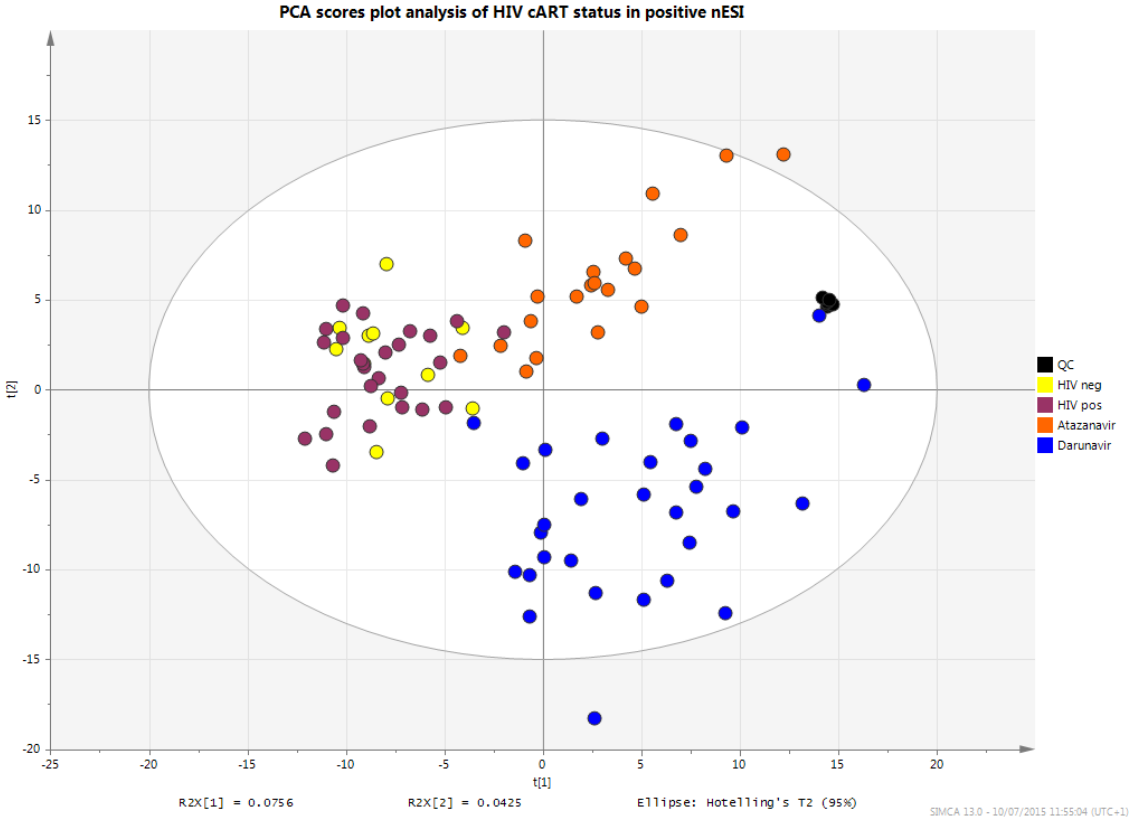


Figure 6.2: PCA scores plot analysis of HIV status and cART intervention in positive ionization mode

The four patient groups are given in the legend, and patients on the two cART medications also received NRTIs tenofovir and emtricitabine, and the ritonavir protease inhibitor. No discrimination was detected between the HIV positive and negative patients. However, both cART groups discriminate from the cART naïve patients on the first component and cART groups cluster separately from each other on the second component.

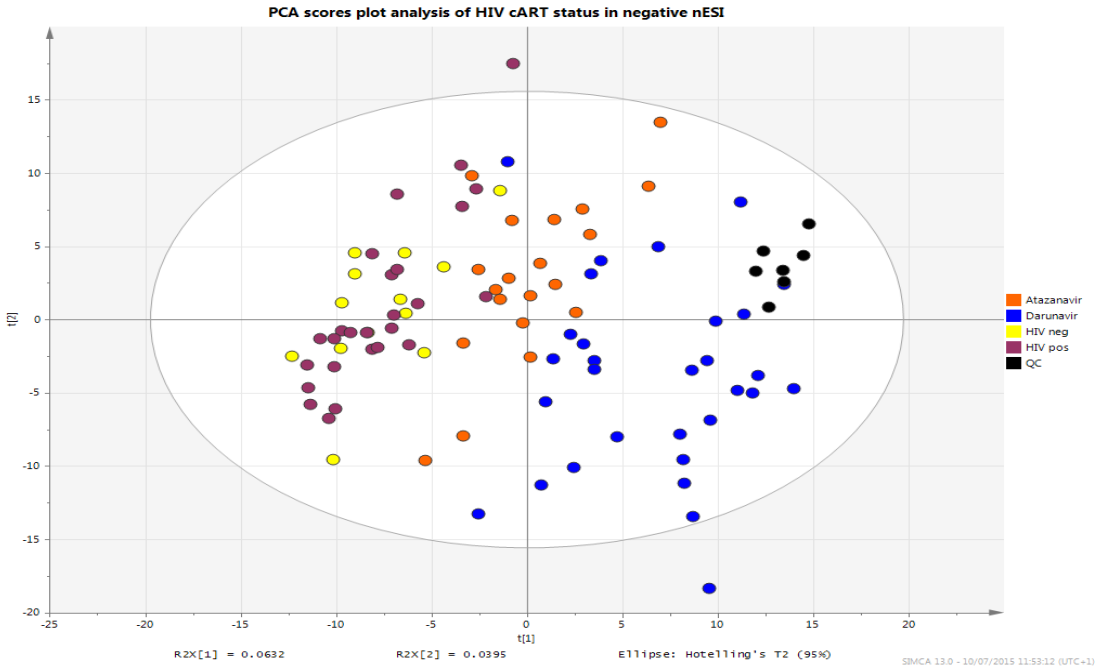


Figure 6.3: PCA scores plot analysis of HIV status and cART intervention in negative ionization mode

The four patient groups are given in the legend. No discrimination was observed between the HIV positive and negative patients. However, both cART groups discriminate from the cART naïve patients on the first component and cART groups cluster separately from each other also on the first component.

In order to identify discriminating markers between cART with cART naïve patients, S-plots from OPLS-DA models were constructed for Atazanavir versus cART naïve HIV positive patients and for Darunavir versus cART naïve HIV positive patients (see example Figures 6.4 and 6.5 respectively for positive nESI mode data sets). The majority of the discrimination observed in the PCA scores plots is driven by ions associated with the cART drugs and their metabolites. This is evident from the large amount of markers present at the top right of both S-plots which are markers of darunavir, atazanavir, ritonavir or their metabolites. In total > 70 ions were positively identified as either the parent compounds, metabolites of cART or their associated fragments all of which are detailed in Table 6.3. The identified metabolites given in Table 6.3, are supported by significant fragmentation patterns which were compared to reported fragments of genuine standards in rodent urine and human blood product and hepatocytes studies (Lin et al., 2013, ter Heine et al., 2009, Vermeir et al., 2009, Zheng et al., 2014). Unlike the PIs, only parent compounds for NRTIs tenofovir or emtricitabine were detected. This is due to the minimal metabolism these drugs experience in the body, and the majority of the drug is excreted as the parent compound (Kearney et al., 2004, Gallant and Deresinski, 2003). All but one of these compounds, Darunavir M4, are highly prevalent in the positive mode dataset, and as such only S-plots from the positive nESI mode are shown. This is the first time these metabolites have been reported in a human urine metabolomic study. The table does not

include any of the significant number of adducts of the molecular ions also identified from the S-plot analysis; some of these are however labelled in Figures 6.4 and 6.5. When adducts and dimers are taken into account there are >100 ions associated with the 3 PIs taken in this study.

The ability to detect these metabolites of the 3 PI drugs highlights the sensitivity of the nLC-nESI-TOFMS and SPE sample preparation technique, as less than 20% of the total parent compound and its metabolites are excreted via the urine (see Table 6.1) (Rittweger, 2007, Le Tiec et al., 2005, Vermeir et al., 2009). The main route of excretion for these compounds is via biliary excretion (Rittweger, 2007, Le Tiec et al., 2005). By using conventional sample preparation and LC-MS analysis, several of these metabolites such as ritonavir metabolites M1 and M2 were unlikely to have been detected. The ability to detect the parent drug and a wide range of its metabolites suggest that in future studies, links between pharmacogenetics, the role of genetics in an individual's drug response, and metabolomics effects could be investigated. This may lead to a more personalised approach to pharmaceutical interventions for a wide range of conditions where individual variation in the extent of drug metabolism and persistence may lead to different treatment options being implemented (Song et al., 2012, Trupp et al., 2012, Agúndez et al., 2009).

In addition to investigating the metabolism and/or the effect of pharmaceuticals on the urinary metabolome, it is also possible to identify any individuals who either have not taken the medication, or are particularly high metabolisers. Indeed, this is evident in patient 284 from the darunavir group where neither darunavir nor any of its metabolites or other cART drugs were detected in their urine sample (Figure 6.6). The BPI chromatograms of patient 284 and 302 (Figure 6.7) clearly show that both darunavir and ritonavir are not present in the urine sample of patient 284. The fact that no other cART drug or metabolites are detected makes it unlikely that the patient took their medication prior to providing the urine sample.

The metabolites detected at the other end of the OPLS-DA S-plots such as the highlighted carnitine suggest that cART intervention has led to changes in the endogenous metabolome. These findings are detailed and their biological implications are discussed in the following section.

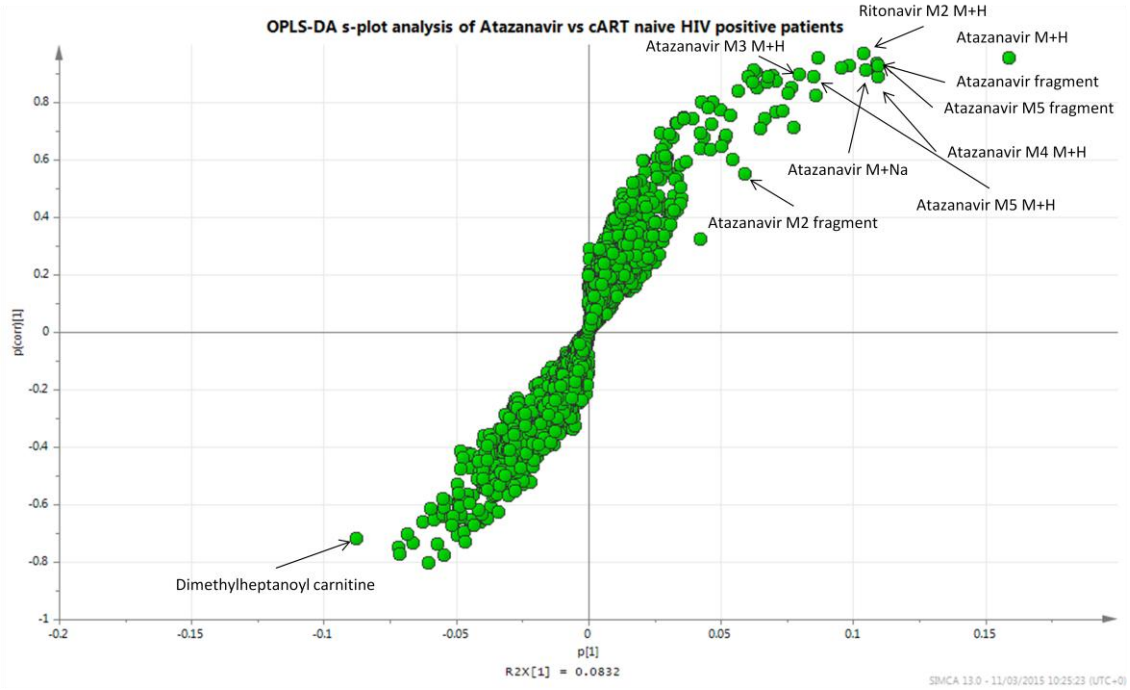


Figure 6.4: OPLS-DA S-plot analysis for patients taking atazanavir therapy compared to cART naive HIV positive patients

Samples were analysed in positive nESI mode. Many of the main discriminating markers are metabolites of the cART therapy. Many metabolites have only been described in non-primate species or cell lines prior to this study.

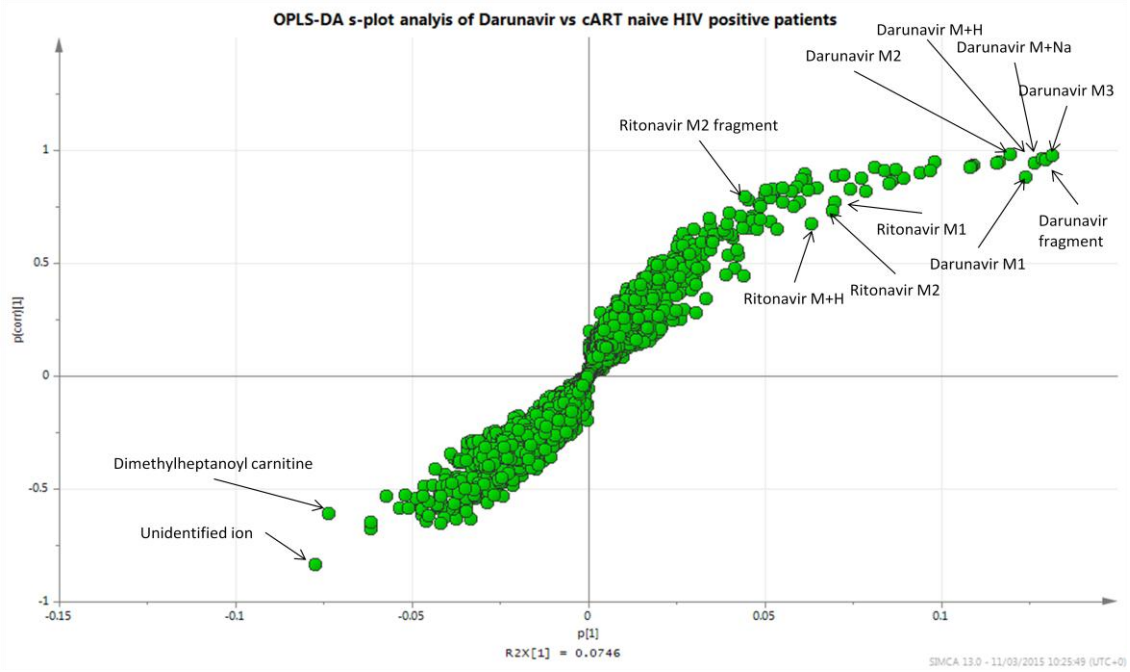


Figure 6.5: OPLS-DA S-plot analysis for patients taking darunavir therapy compared to cART naive HIV positive patients

Samples were analysed in positive ESI mode. Many of the main discriminating markers are metabolites of the cART therapy.

Table 6.3: Detailed identity of cART and related metabolites identified by comparison of cART patient metabolomes with that of cART naive patients

Metabolite	Metabolite ID	Rt	Exp. mass	Thr. mass	Formula	Fragments
Ritonavir	Parent compound	24.09	721.3186	721.3191	$C_{37}H_{48}N_6O_5S_2$	551.2606, 533.2429, 426.1936, 296.1448, 268.1936, 197.0763, 167.086
Ritonavir M1	Hydroxylation	17.96	737.3163	737.3155	$C_{37}H_{48}N_6O_6S_2$	719.3021, 426.1872, 312.1411, 284.1420
Ritonavir M2	Deacylation	17.28	580.3333	580.3321	$C_{32}H_{45}N_5O_3S$	410.2411, 295.1519, 285.1958, 268.162, 250.1606, 171.1004
Darunavir	Parent compound	18.41	548.2439	548.2430	$C_{27}H_{37}N_3O_7S$	392.2005, 241.1039, 202.1623, 156.0143, 113.0626
Darunavir M1	Carbamate hydrolysis	12.00	392.2014	392.2008	$C_{20}H_{29}N_3O_3S$	241.1021, 156.0147
Darunavir M2	Carbamate hydrolysis and hydroxylation	10.09	408.1971	408.1957	$C_{20}H_{29}N_3O_4S$	390.1872, 156.0143
Darunavir M3	Glucuronide of M2	9.67	584.2279	584.2278	$C_{26}H_{37}N_3O_{10}S$	348.0390, 257.0811, 172.0091
Darunavir M4	Glucuronide of parent	14.11	722.2585*	722.2595*	$C_{33}H_{45}N_3O_{13}S$	175.0238, 157.0137, 113.0239
Atazanavir	Parent compound	19.43	705.3964	705.3976	$C_{38}H_{52}N_6O_7$	534.3081, 335.1984, 168.034, 144.1045, 120.0834
Atazanavir M1	Deacylation	14.49	538.3255	538.3241	$C_{26}H_{43}N_5O_7$	367.2264
Atazanavir M2	Carbamate hydrolysis	12.59	647.3929	647.3921	$C_{36}H_{50}N_6O_5$	534.3082, 192.1040, 168.0835
Atazanavir M3	Carbamate hydrolysis	13.06	647.3912	647.3921	$C_{36}H_{50}N_6O_5$	534.3082, 335.1956 168.0835
Atazanavir M4	Hydroxylation	15.41	721.3910	721.3925	$C_{38}H_{52}N_6O_8$	534.3086, 351.1931, 168.0836
Atazanavir M5	Keto metabolite	21.18	719.3770	719.3768	$C_{38}H_{50}N_6O_8$	701.3637, 530.2915, 363.1786, 197.0829, 168.0824
Tenofovir	Parent compound	5.27	288.0884	288.0862	$C_9H_{14}N_5O_4P$	176.0958, 159.0730
Emtricitabine	Parent compound	5.16	248.0522	248.0505	$C_8H_{10}FN_3O_3S$	130.0507, 114.0944

*Refers to m/z of M-H ion, all other ions reported are M+H. RT refers to retention time and Exp. And Thr. mass to experimental and theoretical masses respectively. Details of fragmentation and order of the relative retention time for ritonavir metabolites, atazanavir metabolites and darunavir metabolites were acquired from pharmacokinetic studies in mice and human blood products and hepatocytes (Vermeir et al., 2009, Lin et al., 2013, ter Heine et al., 2009, Zheng et al., 2014)

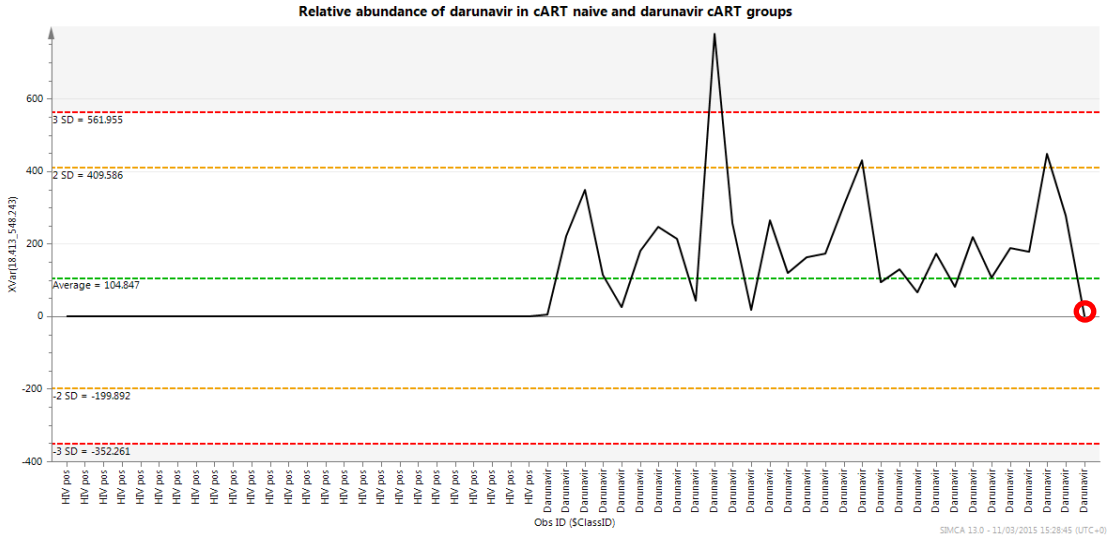


Figure 6.6: Loadings plot for darunavir parent compound

It is evident that one individual (red circle) did not take any cART medication on the day of providing the urine sample. The same pattern is present for metabolites of darunavir or any other cART pharmaceutical they were prescribed.

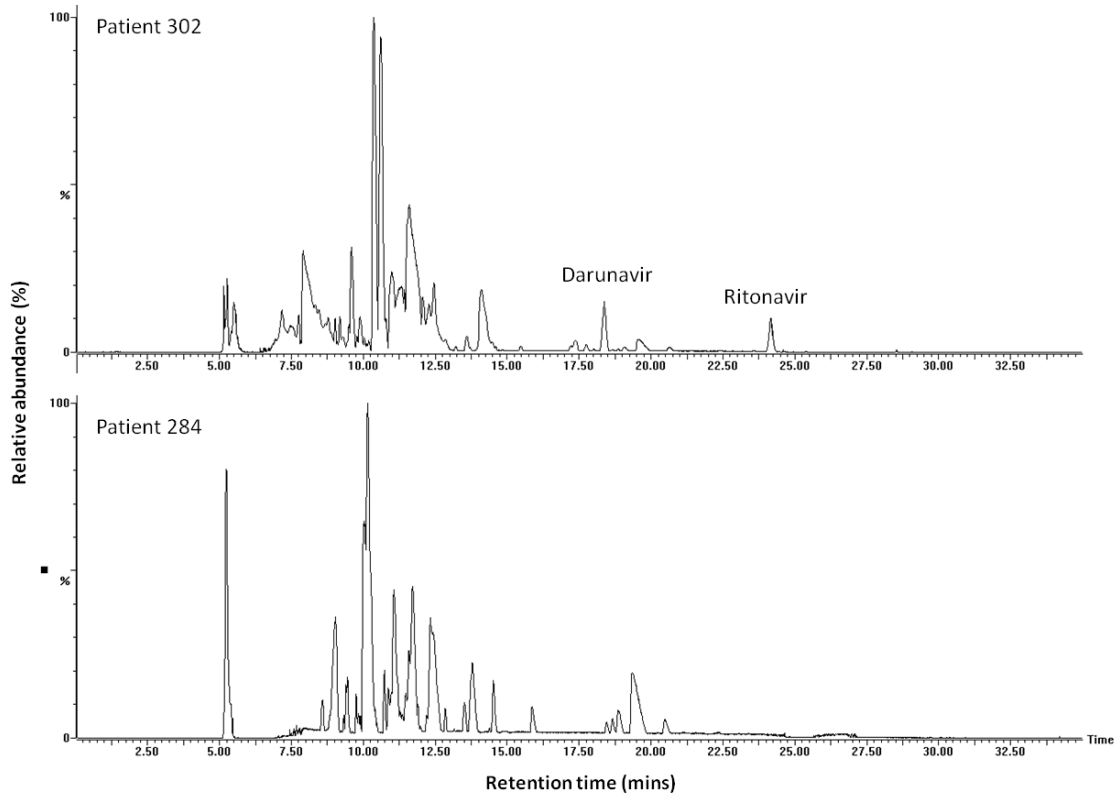


Figure 6.7: Positive ESI BPI chromatogram of urine extracts obtained from patients 302 and 284 with darunavir and ritonavir peaks highlighted in patient 302

The missing peaks corresponding to both darunavir and ritonavir in patient 284 compared to 302 clearly indicate that the cART interventions were not taken in patient 284.

6.4.4 Effect of cART intervention on the endogenously derived metabolome

Following the exclusion of >100 ions associated with the PIs and their metabolites, including fragment, adduct and dimer ions the extreme end of the S-plots were still dominated with more ions associated with cART drugs. Many of these proved to be additional nESI artefacts of the drug molecules such as further adducts, dimers and trimers. Many more were unidentified but shared a common peak shape and retention time with the identified parent compound or their metabolites. This suggests that other less common adducts or fragments were formed which had not been previously reported in the literature. Due to the high abundance of metabolites associated with cART therapy, further analysis concentrated at the other end of the S-plots, i.e. mainly on metabolites that have been decreased due to cART intervention. These metabolite identities have been given in Table 6.4. and their mean normalised peak areas in the patient groups are shown in Table 6.5. Using the D'Agostino-Pearson omnibus K2 test for normality, these metabolite were found to be non-normally distributed and as such a non-parametric Kruskal-Wallis one-way ANOVA was carried out to determine p-values. Discriminating metabolites were only considered significant with a p-value of 0.01 or lower rather than the more commonly utilised p-value of 0.05 to reduce the incidence of type 1 and 2 statistical errors (false positive and negative respectively). To determine the predictability of these discriminating markers, Receiver-Operator Characteristic (ROC) curves were constructed using Prism Graphpad software. ROC curves plot the false positive rate (specificity) against the true positive rate (sensitivity), and the area under the curve (AUC) is the probability that the obtained result is a true result (Broadhurst and Kell, 2006, Seli et al., 2011, Chen et al., 2013a). If the AUC is 0.5, the lowest possible value, there is no significant variation in distribution between the case sample and control e.g. between cART naïve and atazanvir or darunavir groups. The closer the AUC is to 1 the greater the predictability of the variable, and this is also reflected in a steep gradient of the ROC curve. In metabolomics an ROC >0.8 is considered to be good and >0.9 excellent (Broadhurst and Kell, 2006). All the AUC curves for the 11 discriminating markers are included in Table 6.4 with Figures 6.8a and 6.8b showing example ROC curves for 5-deoxy-5(methylthio)adenosine in the cART naïve HIV positive patients versus the darunavir group and pregnenolone sulfate in the cART naïve HIV positive versus the atazanavir group respectively.

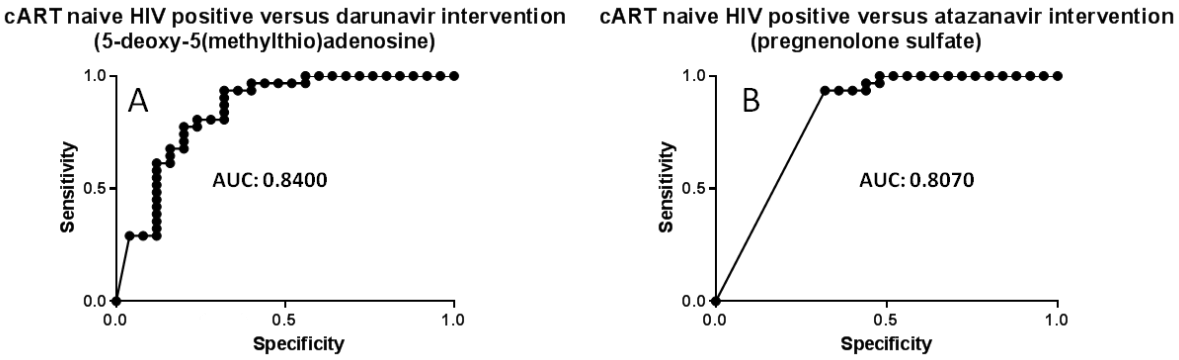


Figure 6.8: a) ROC curve for 5-deoxy-5(methylthio)adenosine in cART naive HIV positive patients against darunavir patients and b) ROC curve of pregnenolone sulfate for cART naive HIV positive versus atazanavir intervention groups

Both ROC curves and AUCs offer good predictability for these markers between treatment group and cART naïve HIV positive groups.

All ROC AUC values for markers down-regulated in either cART treatment are above 0.8, other than cholic acid for darunavir vs cART naïve HIV positive combination, suggesting the predictability is good. The one marker, tetra-peptide, which was detected as up-regulated in the cART groups was just above 0.75 suggesting that for this marker the predictability was lower than for the down-regulated markers.

Due to the post analysis normalisation in this study being to MSTS rather than MSTUS there is concern that the highly abundant metabolites of the cART drugs skew the normalisation in the data sets from cART patient groups and thus artificially suppress endogenous metabolites. However, this is unlikely to be the case in the negative nESI data set due to these drugs being ionized mainly in positive ESI, and here > 100 ions associated with the cART drugs were removed and the data re-normalised. In addition, statistical analysis of the non-normalised data sets was carried out and resulted in the same p-values (all < 0.01) for the identified discriminatory metabolites as that of the normalised dataset. This indicates that any effect of the normalisation method on the detection of discriminatory metabolites was minimal.

Several of the discriminating metabolites identified in this study are of the same compound classes identified in a recent plasma metabolomic analysis of cART intervention (Cassol et al., 2013). In the Cassol et al., (2013) study, the plasma concentrations of taurocholic acid, taurodeoxycholic acid and glycocholic acid were increased in addition to acylcarnitines. In addition, a targeted profiling of bile acid concentrations in urine of HIV positive patients taking PIs, versus HIV negative patients found lithocholic acid and taurocholic acid levels to be significantly increased in the plasma of the HIV positive patients (McRae et al., 2010). However, concentrations of other bile acids, chenodeoxycholic acid, deoxycholic acid, colic

acid and ursodeoxycholic acid were not disrupted in either patient group (McRae et al., 2010). It was suggested that increases in bile acid concentrations may be a result of increase plasma cholesterol, a bile acid precursor, however no such correlation was found (McRae et al., 2010). A further suggested mechanism for increased bile acid levels in the plasma is the inhibitions of bile acid transport into and out of hepatocytes. The uptake of bile acids from the blood into hepatocytes is controlled by membrane transporters. These membrane transporters are known as organic anion transporting polypeptides (OATP). OATPs are typically required for the uptake of xenobiotics which undergo phase 1 and 2 metabolism within the hepatocytes (Kalliokoski and Niemi, 2009). However, these transporters also have a number of endogenous ligands such as both free and conjugated bile acids (Kalliokoski and Niemi, 2009). Two OATP subtypes, OATP2B1 and OATP1B1, are known to be inhibited by PIs including atazanavir, darunavir and ritonavir, although the latter has the greater inhibitory effect of the three (Annaert et al., 2010, Griffin et al., 2011). A reduced uptake of bile acids from the plasma as a result of PI-mediated OATP inhibition would result in increased plasma bile acid concentrations. In addition, apical membrane transporters which export bile acids from the hepatocytes into the bile are also inhibited by PI (McRae et al., 2006). A number of PIs including ritonavir have also been found to inhibit the Bile Salt Export Pump (BSEP) also known as ATP binding cassette transporter B11 (ABCB11). ABCB11 which facilitates the export of both free and conjugated bile acids into the bile (Griffin et al., 2013, Morgan et al., 2010, McRae et al., 2006). Reduced export of bile acids into the bile and build up in the hepatocytes is a cause of hepatotoxicity due to the toxicity of bile acids (Morgan et al., 2010). The inability to excrete bile acids from hepatocytes may result in an increased plasma concentration, or the hepatotoxicity may lead to bile acids leaking from damaged hepatocytes back into the blood.

In the current study bile acid profiles were also found the have been disrupted as a result of PI containing cART intervention (Table 6.4, 6.5). However, unlike the previously discussed reports this work discovered that several bile acids had decreased urinary concentrations. In total 7 free and conjugated bile acids and 1 bile alcohol were detected in urine extracts (Table 6.6). These encompass the primary bile acids, cholic acid, taurocholic acid, glycocholic acid and glycochendeoxycholic acid. The primary bile acid taurochendeoxycholic acid was not detected in these samples. In addition, secondary bile acids, which are a result of C7 dehydroxylation by gut microflora (Lefebvre et al., 2009), were detected and these included deoxycholic acid and its taurine conjugate taurodeoxycholic acid. The other major bile acid formed via microbial metabolism, lithocholic acid, was not detected. The metabolic pathway of bile acid synthesis is shown in Figure 6.9. Of the detected bile acids only cholic acid ($p<0.01$), its glucuronide

conjugate ($p<0.0001$) and glycocholic acid ($p<0.0001$) decreased significantly in the cART treatment groups. Furthermore norcholestanehexol glucuronide was also found to be significantly lower ($p<0.0001$) in the cART treatment group. This bile alcohol is a minor metabolic end point of cholesterol metabolism and its disruption may reflect the changes in the other bile acid profiles. While the changes in bile acid profiles are different to the findings of Cassol et al., (2013) and McRae (2010), the fact that the carnitine metabolite was correlated to the bile acid disruption as it was in the Cassol study implies that fatty acid β oxidation has been disrupted too. This is a potential result of the bile acids themselves being nuclear receptor ligands for receptors such as Peroxisome Proliferator-Activated Receptors, Farnesoid X Receptors, Constitutive Androstane Receptors and Pregnane X Receptors which regulate lipid metabolism in addition to inflammation and the innate immune response (Cassol et al., 2013, Chiang, 2009). In cases of hepatotoxicity, increased plasma bile acid concentrations are usually reflected in the urine (Bathena et al., 2015b, Bathena et al., 2015a). Hence, it was unlikely that the decreased concentrations of certain bile acids observed in the cART exposed patients in the current study were a result of hepatotoxicity. Instead the role of bile acid transporters such as the OATPs and ABCs could be implicated in the decreased urinary bile acid profiles. It is known that these receptors are also present in the kidney (Aleksunes et al., 2008, Huls et al., 2006). This suggests that they may be involved in the excretion of bile acids into the urine, and thus if they are inhibited by PIs it is possible for bile acid concentrations to be decreased in the urine. However, in the current study only a subset of bile acids detected in the urine extracts were disrupted in cART exposed patients which implies that these transporters may have different affinities for the various bile acids. However, it is also noteworthy that the disrupted bile acids in this study were cholic acid or its metabolites, whereas concentrations of chenodeoxycholic acid and its metabolites were not observed to vary between patient groups. These bile acids can be formed by different metabolic pathways (Figure 6.9). Chendeoxycholic acid can be produced via an alternative pathway (Figure 6.10) where it does not rely of CYP7A1 or CYP8B1 enzyme activities. The classical pathway which is the only route for cholic acid production is limited to the liver whereas the alternative pathway for chendeoxycholic acid can occur in the mitochondria of most tissues in the body and may contribute a large proportion of formation of this bile acid (Chiang, 2009). As such, any disruption to CYP7A1 and CYP8B1 will adversely affect cholic acid metabolism whereas chenodeoxcycholic acid metabolism may be affected to a lesser extent. So an alternate hypothesis to explain the decreased concentrations of cholic acid metabolites in urine extracts of samples from cART patients maybe the inhibition of selected CYP enzymes crucial to cholic acid biosynthesis.

Interestingly the fold decreases in bile acid concentrations are much greater (up to 10 times) in the atazanavir group than for darunavir group (Table 6.5). The cause of this is not clear and has not been previously reported as no other metabolomic study has directly compared two different PIs. However, the difference in fold change of cholic acid glucuronide between the two cART regimes may in part be explained by the inhibition of uridine 5'-disphospho-glucuronosyltransferases which atazanavir is known to inhibit whereas no such inhibition has been observed for darunavir (Zhang et al., 2005, Martinez et al., 2014). Inhibition of this class of enzyme prevents the glucuronidation of metabolites and may explain the decreased concentration of the glucuronide conjugate with atazanavir medication. Interestingly using conventional sample preparation techniques and UHPLC-ESI TOFMS analyses, cholic acid glucuronide would not have been detected in urine samples in this study (see Chapter 3).

In the Cassol et al., (2013) study, plasma concentrations of pregnenolone, androsterone and dehydroepiandrosterone sulfates were found to be significantly decreased in the cART patients relative to the healthy controls. While in this study only pregnenolone was detected as being significantly decreased ($P<0.01$) in cART patients, androsterone sulfate levels were also typically lowered in cART groups. However the decrease in concentrations were not significant and this was possibly due to the fact that it appeared to co-elute with an additional androgen sulfate making the peak picking process more complex. Interestingly, the HIV negative patients had the highest level of pregnenolone sulfate and although it was not significantly greater than the cART naïve HIV positive patients it does correlate with the findings of Cassol et al., (2013). In the 2013 study, the levels of steroid sulfates were found to be inversely proportional to that of interferon and interleukin plasma protein markers of pathogen infection. It was found that interferon down regulates sterol synthesis thus offering a potential explanation for the down regulated steroid sulfates (Cassol et al., 2013). Further work using targeted LC-MSMS analyses to quantify steroid profiles in urine extracts is necessary to determine how profiles of these metabolites change in response to viral load and cART medication.

In the current study a nucleoside, 5-deoxy-5(methylthio)adenosine, was elevated in the cART naïve HIV positive patients (Table 6.4). This metabolite has previously been detected, using targeted approaches, at elevated levels in the urine of immunocompromised children (Mills and Mills, 1985, Mills et al., 1985). This variation may be a result of its role in the regulation of apoptosis through the inhibition of protein carboxymethyltransferase which is involved in protein methylation, which in turn can modulate cell signalling and protein expression, although the exact mechanism behind its role in apoptosis is still unclear (Avila et al., 2004, Lee and Cho, 1998). This role in apoptosis is particularly relevant in HIV as a major cause of CD4+

cell loss is due to an increased rate of apoptosis (Levy, 1993, Pantaleo et al., 1993a). In addition, 5-deoxy-5(methylthio)adenosine is involved in the activation of lymphocytes and thus may also be elevated as part of the immune response to HIV infection (Avila et al., 2004). This is supported by the fact that in this study the highest concentration of 5-deoxy-5(methylthio)adenosine was detected in the cART naïve HIV positive patients (Table 6.5). However, differences in detected 5-deoxy-5(methylthio)adenosine concentrations between cART naïve HIV negative and HIV positive patients were not significant. Despite this, it suggests that 5-deoxy-5(methylthio)adenosine may be a marker of a compromised immune system. The significant reduction in urinary concentration of 5-deoxy-5(methylthio)adenosine, could be an indication of the efficacy of cART intervention in these patients as all of them show good viremic control (Table 6.2). Further investigation of 5-deoxy-5(methylthio)adenosine is required due it potentially providing a robust urinary marker of immune function and cART efficacy.

The link between the dipeptide and tetrapeptide markers and HIV/cART patients is not clear, although decreased urinary concentrations of leucine-proline has been observed in both urine and plasma of rat models of atherosclerosis. Atherosclerosis, like HIV, is a long term inflammatory condition linked to disrupted lipid metabolism (Zhang et al., 2009) and in the current study, levels of leucine-proline decreased in urine of cART patients which also showed disruptions in lipid (carnitine and bile acid) metabolism. In addition, urinary metabolomic approaches have been shown that leucine-proline to be increased in patients with colorectal cancer and bladder cancer (Wang et al., 2010b, Huang et al., 2011). In the colorectal cancer study this increase was linked to a possible increased rate of protein catabolism (Wang et al., 2010b). This raises the question as to whether there has been changes in protein catabolism in cART experienced patients which my account for the lowered levels of leucine-proline in their urine.

In the current study a tetrapeptide was found to be increased in the cART groups, however the link between cART treatment and tetrapeptide levels is unclear and requires further investigation. These findings indicate that small peptides may play a significant role in cART mediated effects on metabolomic profiles which to date is under appreciated. Future peptide orientated investigations may further elucidate mechanisms of HIV infection and treatment.

Despite the significantly increase plasma bilirubin levels (Table 6.1) no markers associated with bilirubin or any of its metabolites, bilirubin sulfate, urobilin and urobilinogen were detected in this study. Increased urine levels may have been expected given the increased plasma

concentrations of bilirubin. However, this is possible a result of haem metabolite excretion being primarily via the faeces (circa. 95%) meaning urinary concentrations may be too low to detect (Berk et al., 1969).

It should be recognised that the identities of these discriminating metabolites are putative based upon accurate mass measurement, isotope fit and comparison of fragmentation to MSMS databases and genuine standards (bile acids) where possible. Further targeted analysis is required for metabolite identity confirmation using multiple reaction monitoring (MRM) by comparing genuine standards to the urine metabolites (Dunn et al., 2013, Sumner et al., 2007). In addition, MRM would enable a quantitative analysis of these metabolite concentrations.

Table 6.4: Putative identification of significantly (p<0.01) discriminating markers between cART naive and cART positive HIV patients

ROC atz vs HIV +ve (fold change)	ROC drn vs HIV +ve (fold change)	Rt	Exp. mass	Thr. Mass	Formula of ion	Putative identification	Fragments
Decreased in cART patients							
Positive ESI							
0.8120 (-3.6)	0.8400 (-4.2)	20.91	304.287	304.2852	$C_{17}H_{37}NO_3$	Unidentified metabolite*	none detected
0.9280 (-2.4)	0.8748 (-2.1)	12.57	247.167	247.1658	$C_{11}H_{20}N_2O_3$ (+H_2O)	Leucine-proline +H_2O	229.1551, 183.1473
0.8880 (-10.3)	0.8400 (-5.5)	5.41	298.099	298.0974	$C_{11}H_{15}N_5O_3S$	5-deoxy-5(methylthio)adenosine	136.066, 145.031, 119.038
0.8780 (-8.8)	0.8443 (-3.5)	13.19	302.2342	302.2331	$C_{16}H_{31}NO_4$	2,6 dimethylheptanoylcarnitine	85.0302, 243.1611
Negative ESI							
0.9630 (-34.4)	0.9090 (-10.2)	13.87	583.3124	583.3124	$C_{30}H_{48}O_{11}$	Cholic acid glucuronide^{+}	343.2650, 175.0248, 157.1234
0.8019 (-5.4)	0.7500(-3.6)	18.60	407.2792	407.2797	$C_{24}H_{40}O_5$	Cholic acid	345.2794, 343.2629, 327.2643
0.9370 (-30.1)	0.8645 (-3.6)	12.54	464.3014	464.3012	$C_{26}H_{43}NO_6$	Glycocholic acid*	402.2981, 400.2866, 384.292
0.9800 (-3.3)	0.9510 (-2.9)	12.32	629.3533	629.3537	$C_{32}H_{54}O_{12}$	Norcholestanehexol glucuronide*	175.0237, 113.024
0.8070 (-2.5)	0.8019 (-2.4)	16.57	397.2051	397.2049	$C_{21}H_{34}O_5S$	Pregnenolone sulfate	None detected
0.8310 (-3.6)	0.8600 (-4.2)	9.92	319.644	Unknown	Unknown	Unidentified metabolite	None detected
Increased in cART patients							
Negative ESI							
0.7540 (+15.3)	0.7548 (+7.4)	12.48	539.2694	539.2690	$C_{21}H_{34}N_{10}O_7$	Gln Arg His Thr tetrapeptide	None detected

atz: atazanavir, drn: darunavir, Exp.mass: experimental mass, Thr. Mass: theoretical mass. * metabolites also significantly higher (p<0.01) in HIV negative group versus both cART groups, $^{+}$ metabolites also significantly greater in HIV negative group versus Atazanavir group.

Table 6.5: Normalised mean areas of discriminating endogenous metabolites

Putative metabolite identity	Mean ± standard deviation of normalised metabolite concentration			
	cART naïve HIV -ve	cART naïve HIV +ve	Atazanavir	Darunavir
Leucine-proline +H_2O	23.4±8.7	24.5±11.6	10.2±4.6	11.6±6.0
5-deoxy-5(methylthio)adenosine	5.1±5.5	8.4±7.0	0.8±0.5	1.5±1.8
2,6 dimethylheptanoylcarnitine	514.4±711.7	277.9±322.6	31.8±41.8	78.6±183.2
Cholic acid glucuronide	4.4±5.9	10.8±11.9	0.3±0.3	1.1±2.4
Cholic acid	2.1±4.1	5.4±9.4	1.0±1.9	1.5±2.6
Glycocholic acid	6.3±6.9	6.4±3.7	0.2±0.1	1.8±5.6
Norcholestanehexol glucuronide	7.0±6.9	8.4±3.6	2.5±0.9	2.9±3.4
Pregnenolone sulfate	0.7±0.7	0.9±0.6	0.4±0.1	0.4±0.1
Unidentified metabolite	8.3±13.6	11.8±8.7	0.7±0.5	0.7±1.2
Gln Arg His Thr tetrapeptide	0.3±0.3	0.3±0.2	3.9±8.7	1.9±3.1

Table 6.6: Detailed summary of bile acids detected in the urine of cART naïve and experience patients

Bile acid	Rt	Exp. Mass	Thr Mass.	Formula	Fragments	HIV +ve vs atazanavir	HIV +ve vs darunavir
Cholic acid	18.6	407.2792	407.2797	$C_{24}H_{40}O_5$	345.2794, 343.2629, 327.2643	p<0.01	p<0.01
Cholic acid glucuronide	13.87	583.3124	583.3124	$C_{30}H_{48}O_{11}$	343.2650, 175.0248, 157.1234	p<0.0001	p<0.0001
Taurocholic acid	12.83	514.2823	514.2838	$C_{26}H_{45}NO_7S$	124.0067, 106.9804	ns	ns
Glycocholic acid	12.54	464.3014	464.3012	$C_{26}H_{43}NO_6$	402.2981, 400.2866, 384.292	p<0.0001	p<0.0001
Deoxycholic acid	24.38	391.2832	391.2848	$C_{24}H_{40}O_4$	345.2771, 343.2621	ns	ns
Taurodeoxycholic acid	16.62	498.2867	498.2889	$C_{26}H_{45}NO_6S$	124.0067, 106.9807	ns	ns
Glycochenodeoxycholic acid	16.6	448.3055	448.3063	$C_{26}H_{43}NO_5$	nd	ns	ns
Norcholestanehexol glucuronide	12.32	629.3533	629.3537	$C_{32}H_{54}O_{12}$	175.0237, 113.024	p<0.0001	p<0.0001

Exp. Mass = experimental mass, Thr Mass = theoretical mass, HIV +ve refers to patients with HIV who are cART naïve, ns = p>0.05. nd = metabolite signal too weak for detection of fragments

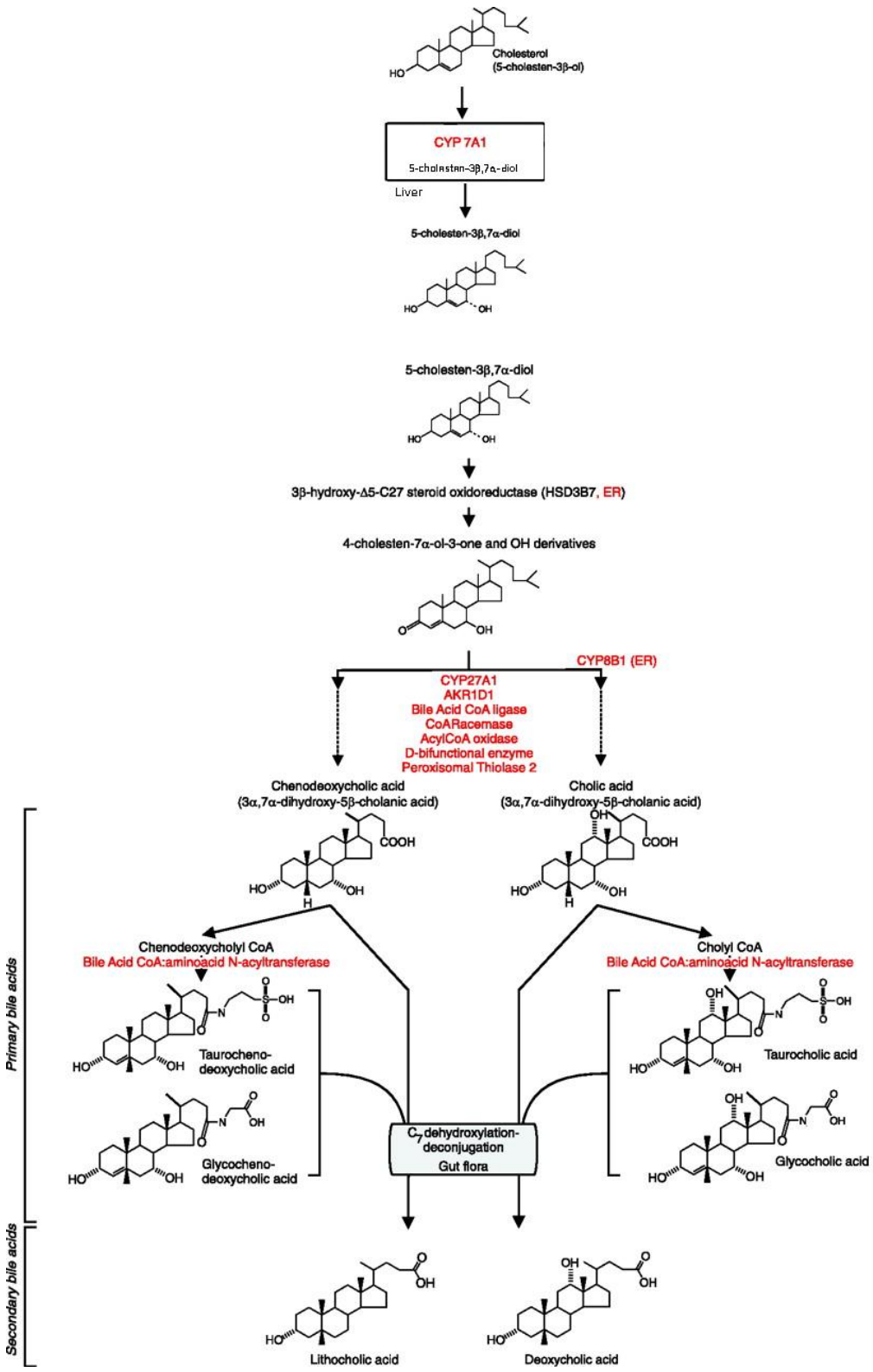


Figure 6.9: Bile acid metabolomic pathway from cholesterol

The discriminating metabolites of the bile acid pathway detected in this study were from the cholic acid portion of this metabolic pathway. (Adapted from "Role of Bile Acids and Bile Acid Receptors in Metabolic Regulation" 2009) (Lefebvre et al., 2009)

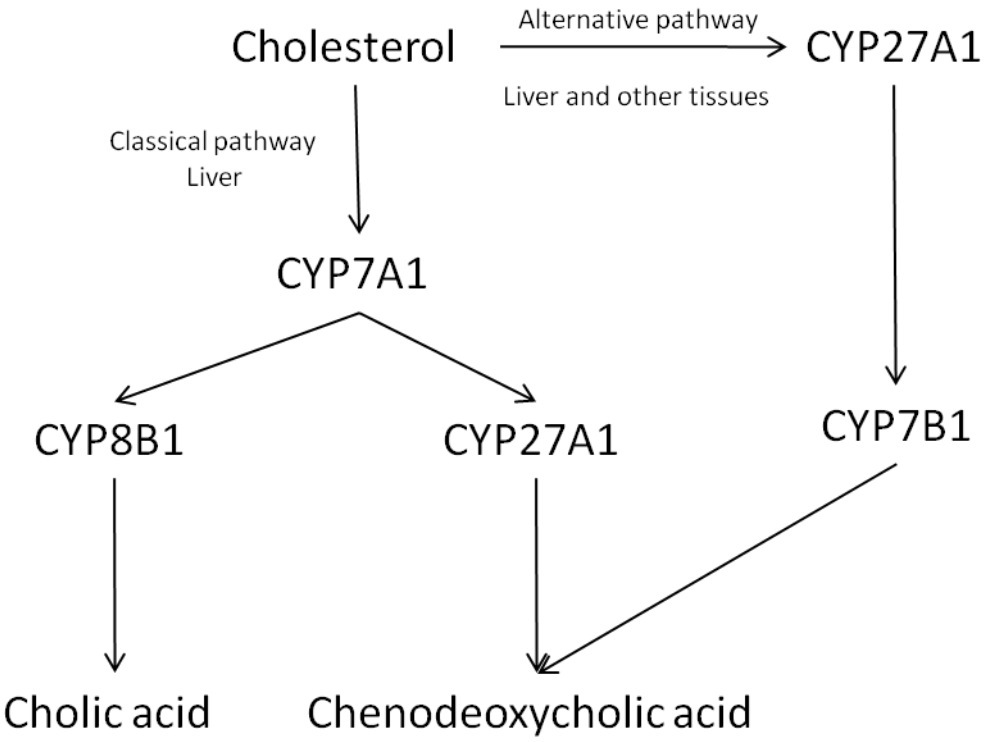


Figure 6.10: Classical and alternative metabolic pathways for the production of primary bile acids

Metabolic disruption to CYP7A1 or CYP8B1 are likely to have a significant effect on the production of cholic acid and its derivatives where as chenodeoxycholic acid production will be affected to a lesser extent. (Adapted from: "Bile acids: regulation and synthesis" 2009) (Chiang, 2009)

6.5 Conclusion

The use of newly developed sample preparation and analytical methodologies for metabolomic analysis of urine has allowed for the first urinary LC-MS metabolomic approach to investigate the metabolomic consequences of cART therapy in HIV positive patients. The current study is also the first to detail the extensive metabolism of PIs in human urine and to detect these metabolites using a non-targeted approach. The ability to detect a wide array of pharmaceutical metabolites in urine extracts of patients opens up the possibility to investigate any differences in drug metabolism for example between high metabolisers and low metabolisers with potential use for future personalised medication. In addition, in this study it allowed confirmation of patients into cART naïve and cART experienced groups. Furthermore, an array of endogenous metabolites was found to distinguish between cART naïve and cART experienced HIV positive patients. These included bile acids which are ligands of several nuclear receptors and aid in the regulation of lipid metabolism, inflammation and immune function. However, unlike previous studies where bile acid concentration in plasma has been observed to increase, in this study selected urinary bile acids were found to decrease in concentration. This is potentially a result of PI inhibition of bile acid transporters and requires further investigation to determine if this is indeed the case. Concentrations of a carnitine metabolite were also decreased in urine of cART experienced patients indicating disruption of mitochondrial lipid β-oxidation. These metabolite changes are the converse of those previously observed in other studies of the effect of cART intervention on the plasma metabolome and warrant further investigation of both matrices to elucidate the metabolomic causes of this disparity. In addition, two further metabolites, leucine-proline and 5-deoxy-5(methylthio)adenosine were found to be reduced in patients receiving cART intervention. This is the first occasion that either have been associated with HIV cART intervention, although 5-deoxy-5(methylthio)adenosine has previously been reported to be elevated in immunocompromised children. The mechanisms behind both of these metabolites require additional studies to quantify and understand the role they have in HIV infection. The fact that two peptide markers were found to discriminate between sample groups implies that a future peptidomic approach may reveal further markers of HIV cART intervention. The finding of this study indicate that the use of the new analytical methods for method development and LC-MS analysis enable a sensitive approach for allowing a non-invasive analysis of cART therapy using metabolomics analyses of urine extracts.

Chapter 7: General discussion

7.1 Introduction

To date, the field of metabolomics has been limited to the analysis of only the most abundant metabolites in the metabolome. This focus on the most abundant metabolites in itself has led to significant advances in our knowledge of disease and is currently making changes to patient quality of life and treatment. On example of this is the development of the so called 'intelligent knife', where the smoke plume from a cauterising scalpel is analysed using online mass spectrometry. The enables surgeons to identify cancerous tissues during surgery ensuring the entire tumour is removed (Balog et al., 2013). However, low abundance metabolites are not required in this analysis and like many metabolomics studies these metabolites are not detected. This is both a result of neglected sample preparation techniques and the use of less sensitive analytical platforms. In addition, urinary metabolomics suffers from a unique problem posed by variable concentration within and between individuals.

This thesis begins to offer solutions to these issues by introducing SPE sample preparation methodologies to ensure thorough sample clean up and pre-concentration prior to analysis. In addition, new nanoflow-nanospray LC-MS technologies were investigated as a source of highly reproducible and sensitive analyses for metabolomic studies. Furthermore, a novel normalisation methodology was investigated using the pre-analysis correction of urine concentration according to sample osmolality to minimise the detrimental effect of urine concentration and a post-analysis MS signal normalisation to reduce fluctuations in LC-MS performance prior to multivariate statistical analyses. As a result of this work, the first LC-MS urine metabolomic analysis of HIV patients was carried out investigating the metabolomic consequences of HIV infection. This analysis proved to be highly sensitive and capable of detecting metabolites that would not normally have been detected during metabolomic analysis without the newly developed SPE sample preparation techniques and nanoflow-nanospray analysis methodologies. Despite this sensitive analysis it was not possible to ascertain any discriminating metabolites between groups of HIV negative, acute or chronic HIV infection groups. However, a likely cause of this is that NSAIDs were detected in the majority of patients meaning that any inflammatory markers would have been suppressed. This highlights the importance of obtaining an accurate patient history prior to metabolomic analysis to ensure patient samples can be stratified by pharmaceutical intake to avoid confounding the analysis. In addition, the patients in this study were in the early stages of HIV infection before antiretroviral therapy was required suggesting any metabolic effects of

infection may be too subtle to be detected in the urine and that an analysis of the plasma may be better suited for metabolomic analysis of early HIV infection.

A second study investigated more advanced HIV infection and the consequences of two cART intervention regimes on the urinary metabolome compared with cART naïve HIV positive and negative groups. Here the cART intervention groups contained three common pharmaceuticals, the NRTIs tenofovir and emtricitabine and the PI ritonavir. The two groups contained different additional PIs, one group contained darunavir and the other atazanavir. Analysis of the urine metabolomic profiles of these patients highlighted that the cART intervention significantly altered the biochemical nature of urine. Cholic acid and several of its metabolites were discovered to be down regulated in the urine and the proposed mechanisms for this was disruption to the classical bile acid synthesis pathways or inhibition of bile acid transporters in the kidney. This was the first work to report a reduced concentration of bile acids in the urine of patients receiving cART. This opens up further opportunities to enhance our understanding of cART interventions on bile acid metabolism and elimination as previous plasma studies indicated that bile acid levels increased in cART experienced patients (McRae et al., 2010, Cassol et al., 2013). Furthermore, 5-deoxy-5(methylthio)adenosine was detected for the first time as a marker of cART intervention in HIV patients. This metabolite is strongly linked to roles in apoptosis and immune function and may prove to be a strong marker of cART efficacy as urinary concentrations dropped dramatically with cART intervention. This study also highlighted for the first time a potential role of small di- and tetrapeptides in HIV infection and therapy, although further studies are required to elucidate these functions. In addition to endogenous metabolites, the metabolism of cART drugs was a large driver of discrimination between sample groups. This was the first time the extensive metabolism of protease inhibitors was determined using a metabolomics approach.

Despite the advances presented in this thesis, significant challenges still exist in metabolomic analysis, and these may be addressed using either new or more commonly available techniques that have been introduced in the last few years. These are discussed in this chapter in addition to offering potential future metabolomic applications of these new sensitive techniques.

7.2 Sample preparation

As previously discussed in this thesis, sample preparation is frequently overlooked in favour of rapid methodologies allowing for high throughput analysis (Fernández-Peralbo and Luque de Castro, 2012, Alvarez-Sanchez et al., 2010b, Alvarez-Sanchez et al., 2010a). This

thesis introduced both mixed mode anion and mixed mode cation exchange SPE, with a pre-concentration step. The use of this methodology proved to significantly increase the number of metabolites detected (Chapter 3). However, this methodology required two cartridges and was completed off-line making it a time consuming process. Recently Phenomenex (Macclesfield, U.K) have developed a single cartridge with reversed phase, cationic and anion exchange stationary phase thus halving the cost, solvent use and sample prep time (Boisvert et al., 2012). A further option is the use of on-line SPE, this technique has been used in some platforms for several years, but is becoming more common with nano scale platforms in the form of column trapping which is often used in proteomic analysis (Koppen et al., 2014, Rogeberg et al., 2014).

Online-SPE/ column trapping is an automated sample preparation method which offers significantly reduced sample preparation times, solvent usage and potential user error (Rogeberg et al., 2014, Anumol and Snyder, 2015, Koppen et al., 2014). In addition, the solvent evaporation stage is eliminated, and this not only reduces the time taken but eliminates the potential sample loss during this process (Anumol and Snyder, 2015). The addition of online SPE/trapping has been shown to have no negative connotations on mean peak area repeatability or reproducibility (Anumol and Snyder, 2015) or retention time stability (Rogeberg et al., 2014). Although reversed phase trapping and analytical columns are currently the only available options for the current work on the Waters nanoAcquity platform, other groups have investigated various combinations of HILIC, mixed mode and reversed phase chemistries for optimised sample clean up, pre-concentration and chromatography (Rogeberg et al., 2014). In contrast to conventional SPE cartridges which are single use only, online-SPE and trapping columns may be used hundreds of times as they are washable, thus significantly reducing the cost and environmental impact of sample preparation (Anumol and Snyder, 2015).

While much of the column trapping used in nLC-nESI-MS allows low abundance proteins to be concentrated and desalted prior to proteomic analysis (van de Meent and de Jong, 2011, Wang et al., 2007), several studies have recently demonstrated its applicability to small molecule analysis (Berlioz-Barbier et al., 2015, Berlioz-Barbier et al., 2014). In these studies column trapping was utilised in the analysis of pharmaceutical contaminants in waste water effluents and mussels. A column trapping methodology was also successfully utilised in a global metabolomic analysis investigating the role of polyphenolic compounds present in olive oil on apoptosis in colon cancer, (Fernández-Arroyo et al., 2012).

The incorporation of an automated online sample preparation method would significantly reduce the time taken to prepare samples for metabolomic analysis. In addition, by automating the process a more reliable robust analysis could be achieved as fewer work up errors are likely to affect the analytical outcome.

7.3 Analytical platforms

7.3.1 Liquid Chromatography

One of the problems encountered in this thesis is the retention of polar metabolites on the nano flow columns. Although much work has been carried out by others in an attempt to solve this problem using HILIC, or aqueous normal phase chemistries, these techniques still require an additional LC-MS run to analyse polar compounds (Zhang et al., 2012b, Spagou et al., 2011, Gika et al., 2008a). A further problem, despite the use of nano flow rates, is that of ion suppression as a result of co-elution of compounds from the analytical column.

One potential solution to this is to incorporate a two-dimensional LC (2D LC) separation prior to MS analysis. 2D LC is known to greatly improve chromatographic resolution, in turn reducing ion suppression, which subsequently increases the signal to noise ratio (Chalcraft and McCarry, 2013, Li et al., 2013b, Genta-Jouve et al., 2014). There are a number of techniques to generate a 2D LC-MS method; one is to use two columns of the same chemistry, while the other is an orthogonal separation where two different chemistries are incorporated (Chalcraft and McCarry, 2013, Jandera, 2012). In both cases both separations occur in series, while in the latter, separation is driven by the differing chemical properties of the analyte interact with two stationary phases (Lu et al., 2008). There are also a number of ways in which the eluent from the first column is introduced to the second. Either they run straight into each other, or fractions are collected online by filling a second sample loop with the eluent of the first column, which is subsequently injected onto the second column. In addition, the 2D separation may be achieved offline, where the fractions from one LC run are saved prior to the second dimension LC analysis (Yin and Xu, 2014, Jandera, 2012). In a comprehensive 2D LC system, all fractions collected from the first column are subsequently analysed by the second column. However, in a process known as “heart cutting”, only selected fractions are analysed again, with the remaining being directed to waste (Jandera, 2012, Yin and Xu, 2014). In non-targeted metabolomics, “heart cutting” 2D LC would be detrimental to the extensive coverage required for the metabolome.

To date, 2D LC has not been widely adopted in the field of metabolomics. One study of the urinary profile of lung cancer patients utilised online 2D LC-MS with both HILIC and RP

column chemistries. This study identified 17 discriminating metabolites encompassing amino acids, bile acids, peptides and organic acids which covered a wide range of polarities, thereby justifying the use of both column chemistries (Yang et al., 2010). A further study by Edwards et al., (2007) employed an offline 2D LC-MS platform where the first dimensional separation was achieved using a cation exchange column. Due to incompatibilities with reversed phase chemistry, the eluent from the first column was collected into 6 fractions, which were dried and reconstituted prior to the second dimensional analysis on a RP column. Compared to conventional 1D RP separation the addition of a second dimension increased the number of metabolites detected by 44%, however the incorporation of an evaporation and reconstitution step significantly increased analysis time (Edwards et al., 2007). In a further metabolomic analysis of human urine, the use of 2D LC-MS enabled 2.5 fold more metabolites to be analysed compared to a single LC-MS analysis. This, as with the Edwards et al (2007) methodology, utilised an offline 2D LC-MS. However in both runs the same column chemistry was utilised, but instead the polarity of the metabolites were altered using a derivatization step following the 1st LC run. Here metabolites were derivatized using ^{12}C-^{13}C dansylation thus changing the retention characteristics of each metabolite. This step alone however took 90 minutes and thus adds a significant amount of time to metabolomic analysis (Guo et al., 2011).

7.3.2 Mass Spectrometry

The use of nanoflow and nano spray platforms undoubtedly increases sensitivity to low abundance metabolites in the metabolome. This is highlighted in the detection of low abundance metabolites in the HIV studies reported in Chapters 5 and 6. However, one major bottleneck in metabolomics is structural elucidation and metabolite identification (Paglia et al., 2014). With more sensitive platforms structural elucidation using NMR is no longer an option, due to the low abundance of such metabolites (Lanucara et al., 2014). A further issue with nanoflow platforms is the relatively long run times (circa 1 hr), and in many cases high throughput analysis is preferred to a more sensitive analysis (May et al., 2015). Although not a new technology, ion mobility (IM) mass spectrometry offers a potential solution to these issues. Ion mobility offers a third dimension in LC-MS separation, by incorporating a drift cell between the ionization step and the entry to the mass analyser. Ion mobility separates ions based upon their collisional cross section (CCS) (Paglia et al., 2014). Here ions are passed into a drift chamber which is filled with an inert drift gas, such as helium or nitrogen, and a weak uniform electric field is applied (Ibrahim et al., Zhou et al., 2014, Shrestha and Vertes, 2014). The electric field is utilized to drag the ions through the drift chamber and the time taken for an ion to pass through the drift chamber is directly proportional to the number of collisions

with the inert gas, which in turn is a direct result of its surface area. This allows IM to separate ions based upon their size and shape before entering a mass analyzer, which then separates ions based upon their mass to charge ratio (Kanu et al., 2008, Zhou et al., 2014, Lanucara et al., 2014, Fenn et al., 2009, Shrestha and Vertes, 2014, Paglia et al., 2014). Separation on CCS allows isobaric, chiral and isomeric ions to be separated within the flow chamber very quickly, and much faster than using a conventional LC methodology (Kanu et al., 2008). In addition, IM can clearly distinguish between chemical classes as they tend to group together in 'drift time' due to their similar molecular structure. For example amino acids, lipids and polypeptides group in distinct groups in drift time (Kanu et al., 2008, May et al., 2015, Fenn et al., 2009, Shrestha and Vertes, 2014). This may help to at least classify the compound class of an unknown metabolite. IM-MS has also been shown to increase the signal to noise ratio by removing unwanted matrix components and reducing noise, which in turn improves the limit of detection of the metabolomic analysis (Fenn et al., 2009, Lanucara et al., 2014, Kanu et al., 2012). The increased separation afforded by IM-MS allows for improved separation when using a shorter LC run time thus allowing for a more high-throughput metabolomic analysis (Lanucara et al., 2014). These attributes of increased structural elucidation, separation of isobaric compounds and improved limits of detections make LC-IM-MS ideal for the analysis of complex matrices such as metabolomics (Kanu et al., 2008). However, to date very few metabolomic analysis has been completed using this methodology. A study by Paglia et al., (2014) investigated the metabolomic profiles of plasma, urine, red blood cells and platelets in 3 different laboratories and demonstrated that LC-IM-MS is highly reproducible, both within and between laboratories (Paglia et al., 2014). Given the benefits of IM it seems likely that it will soon be incorporated into future metabolomic studies to not only increase confidence in metabolite identity, but to provide an increase in sensitivity and potentially reduce analytical run times.

7.4 Data Analysis

One the main rate-limiting steps in metabolomics is the data analysis and metabolite identification (Cui et al., 2008, Creek et al., 2014). This is a problem that is more exaggerated with ever more sensitive methods, as not only are there more peaks but the background noise is also greater. In order to accurately analyse metabolomic workflows, software capable of determining metabolite peaks from noise and MS artefacts are required (Want and Masson, 2011). Ideally, a single peak would correspond to a single ion, however this is very rarely the case due to co-elution, adducts, fragments and dimers/trimers formation, meaning multiple ions are detected for a single peak (Want and Masson, 2011, Katajamaa and Orešič, 2007). In

addition, some peaks are hidden within the noise while others are false peaks generated as a result of noise. This is a particular challenge for the very low abundance metabolites (Yu et al., 2013).

The process of peak picking has several stages before the final peak list is produced. Each software has a unique combination of algorithms for each stage, thus meaning each software package performs differently. The first step is to correct for variation in the baseline, caused by variation in column temperature, build up of contaminants, or mobile phase interference. Following this, a noise reduction step is utilised and this reduces the effect of solvent impurities (Want and Masson, 2011) and, for nanoESI platforms, atmospheric impurities such as oil vapour from the vacuum pumps (Schlosser and Volkmer-Engert, 2003). Next the chromatogram is smoothed, improving peak shape making them easier to detect. Following this the chromatogram is deconvoluted, whereby the ions associated with each peak is given an *m/z* x Rt value. Some software packages such as the firmware Progenesis, or freeware such as XCMS (Smith et al., 2006) and MzMine (Pluskal et al., 2010) are capable of applying multiple ions to a single metabolite (i.e. fragments), whereas others give multiple *m/z* x Rt values for a single compound (Want and Masson, 2011, Katajamaa and Oresic, 2005). Correction for retention time drift due to column aging, matrix effects and temperature variation is also usually required. This can be achieved in one of two ways, either the whole chromatographic profile is shifted to align a single peak such as an internal standard (Want and Masson, 2011), or alternatively a non-linear retention time alignment can be applied where by multiple matching metabolites between chromatograms are used to align retention times, thus providing a more dynamic correction (Want and Masson, 2011, Qi et al., 2012, Zhang et al., 2014, Tautenhahn et al., 2012b, Smith et al., 2006). The final step of the process is to integrate the peaks allowing a relative abundance to be determined (Want and Masson, 2011).

A number of studies have compared different peak picking software packages in an attempt to find the ideal package. One study, investigating uterine cervix cancer compared a firmware Markerview (Sciex, Warrington, UK) with two freeware packages; MZmine and XCMS (Chen et al., 2013a). Analysing QC samples, each software package detected a different number of peaks; XCMS 1651, MZmine 994 and Markerview 2516. This is a result of different packages utilising different peak picking algorithms, and the main problem between the packages arose with low abundance metabolites being confused with background noise. Peaks with an intensity of $>10^4$ counts were generally agreed between packages and the raw data (Chen et al., 2013a). When control samples were compared to the uterine cervix cancerous samples, a total of 14 discriminating markers were detected and identified across the

packages, however only 2 of these markers were common between all the packages (Chen et al., 2013a). In 2014 Coble et al., investigated XCMS (the most cited peak picking software), MZmine and Metalign (Lommen and Kools, 2012). In line with the findings of Chen et al., (2013), a disparity in the numbers of peaks detected was found between different software packages. In total Mzmine detected 38082, Metalign 5107 and XCMS 6204; this study also investigated the percentage of which were false peaks of which 84.7%, 87.2% and 51.8% were incorrectly picked as peaks respectively. In terms of missing peaks, only XCMS did not miss peaks, thus along with the lowest proportion of false peaks, XCMS was determined to be the preferred package on this occasion (Coble and Fraga, 2014). In a more recent analysis Peakview (Sciex, Warrington, UK), Markerview, Metabolitepilot (Sciex, Warrington, UK) and XCMS were compared using bile and urine MS data and a mixture of 84 standards (Rafiei and Sleno, 2015). Neither of the packages were able to peak pick all 84 standards, and in order of the number of standards detected, the packages ranked Markerview, Peakview then jointly XCMS and Metabolitepilot. As with the previous study, a large disparity was noted for the total number of peaks detected for the combined profiles of bile and urine. These totals were 11553, 11708, 2015 and 9879 for Markerview, XCMS, Peakview and Metabolitepilot respectively (Rafiei and Sleno, 2015). It is, however, important to remember that detecting the most number of peaks in itself is not a marker of the quality of peak picking, as this may well be influenced by factors such as background noise. While the quality of the peak picking was not discussed it was noted that XCMS and Markerview provided the quickest peak picking process. However, Peakview and Metabolitepilot allowed for a visual analysis of the extracted ion chromatogram for all peaks detected. Finally both Markerview and XCMS have the ability to detect and remove isotopes meaning fewer ions are peak picked for the same compound (Rafiei and Sleno, 2015).

An additional benefit of XCMS is that the data output can be imported to an excel macro named IDEOM which can be used to reduce the impact of peak shoulders, irreproducible peaks, adducts, dimers/trimers, noise and fragments. The developers of IDEOM claim that these features account for up to 80% of all peaks picked. The remaining peaks may then be compared to either a built-in database or the users own library of standards with retention time data to aid in the identification of metabolites (Creek et al., 2012). However, XCMS online has a limited capacity of 3GB per user initially which is far too small for data files derived from a nanoESI platform even following the data compression required for XCMS upload. The offline version to date only allows for a pair ways analysis, limiting analysis to a

control versus a single case group. In addition, offline XCMS requires additional expertise from the researcher in the R programming language.

What is clear from these studies, is that so far no single peak picking software can be relied upon for accurate peak picking, and any resulting discriminating peaks require extensive manual examination and follow up to assess the veracity of the result (Coble and Fraga, 2014, Chen et al., 2013a). However, a new software package named Progenesis (formerly Transomics) which has found success in proteomics, offers an improved peak picking process for metabolomics. As with XCMS a non-linear peak alignment algorithm is included, but is also available for manual user alignment (Qi et al., 2012, Zhang et al., 2014). Peak picking is achieved using a reference map of all detected peaks which is compared to each chromatogram, and uses a wavelet based peak picking with isotope matching to determine a peaks presence or absence in the chromatogram. This allows for overlapping peaks to be differentiated (Kuharev et al., 2015, Nahnsen et al., 2013). Progenesis also identifies and removes adducts, and can be used to investigate only metabolites with a pre-defined significant difference between sample classes. In addition, Progenesis has a built-in database for metabolite identification, based on accurate mass measurement, and the ability to utilise online databases. In a recent update, MSe data can be used for metabolite identification using additional fragmentation data thus adding a more reliable form of metabolite identification where fragmentation data is available in databases.

Even with these recent developments it is clear that more needs to be done to improve metabolomic peak picking. With the high standards set for retention time and mean peak area repeatability, it appears that this may be lost during the peak picking process.

Once peaks have been successfully picked, discriminating metabolites then need to be identified. Typically this can be done by searching databases such as Metlin (Tautenhahn et al., 2012a), HMDB (Wishart et al., 2013) and MyCompoundID (Li et al., 2013a). However, many compounds are isobaric, and thus these searches may return many compounds of the same mass. Confidence can be increased in metabolite identification by comparison of fragmentation data to pure standards, or MSMS data provided in databases (Tautenhahn et al., 2012a, Creek et al., 2014). Currently the human metabolome database contains very limited MSMS data, however it provides detailed biological importance of each metabolite and predicted concentrations in biofluids (Wishart et al., 2013). Metlin however is another freely available mass spectral database aimed at metabolomic researchers, and containing over 10,000 MSMS mass spectra, and some with 4 different collision energies (Tautenhahn et al.,

2012a). However, both databases contain only a small fraction of the metabolites present in the human metabolome, and as such both need to be expanded. MycompoundID has the additional functionality of identifying compounds based upon fragments and adducts and links to HMDB (Li et al., 2013a). The gold standard for metabolite identification is the comparison of accurate mass, retention times and fragmentation patterns against a pure standard (Creek et al., 2014). This can now be further improved upon for the identification of structural and chiral isomers with the use of ion mobility mass spectrometry.

Once reliable peak picking software and more expansive metabolite databases have been developed, metabolomic workflows would be significantly quicker. In addition the more accurate peak picking process and MSMS metabolite identification would provide a more reliable metabolomic analysis.

7.5 Future metabolomic studies

There are still several areas in which urinary metabolomics have been underutilised and could provide valuable insight. In medical research, various pathologies of the kidney are a significant burden on healthcare, with over 10% of adults developing chronic kidney disease (Levey et al., 2003, El Nahas and Bello, 2005) and 5-7% of hospitalized patients developing kidney disease as result of kidney injury (Chertow et al., 2005). To date, much of the metabolomics research carried out on the kidney has been related to kidney cancers (Huang et al., 2013, Ganti et al., 2012a, Ganti et al., 2012b, Ganti and Weiss, 2011, Kim et al., 2011). These studies have so far only identified disrupted amino acid and carbohydrate metabolism, and offer a non-specific method for detecting kidney cancer (Weiss and Kim, 2012). Other than cancer, kidney pathology of particular interest is acute kidney injury (AKI) as a result of surgical intervention of which two examples stand out; contrast-induced AKI and AKI following cardiac surgery (Solomon and Dauerman, 2010, Rosner and Okusa, 2006). In the case of contrast-induced AKI, contrast solution is administered to patients undergoing angiography and causes kidney injury in up to 25% of angiography patients (Solomon and Dauerman, 2010, Seeliger et al., 2012). Another significant cause of AKI is during coronary artery bypass surgery, during which the kidney suffers from hypoperfusion-induced AKI. This occurs in between 5 and 30% of patients at a cost of >$20,000 each in the U.S.A (Dasta et al., 2008, Chertow et al., 2005, Rosner and Okusa, 2006). In both examples the AKI occurs at the time of surgery, however current methods to diagnose AKI, namely elevated serum creatinine, take several days to become apparent during which the patient has lost significant kidney function (Seeliger et al., 2012, Solomon and Dauerman, 2010, Weiss and Kim, 2012, de Geus et al., 2012). To counter this, an earlier indication of AKI is required to ensure patients receive rapid treatment to avoid

permanent kidney damage, dialysis and mortality (Chertow et al., 2005, de Geus et al., 2012). To date proteomics has identified several protein markers indicative of AKI, however these have proven to be both expensive and unreliable (de Geus et al., 2012). Ideally patients at risk of AKI could be indentified prior to surgery to allow immediate intervention to protect the kidney. To do this large scale sensitive metabolomic studies of these patients are required to identify potential predictive biomarkers of AKI.

In addition to direct medical research, the use of highly sensitive metabolomic analysis offer significant improvements in the field of exposomics. The exposome is the life time exposure of an organism from conception to death and how these exposures may impact upon health and disease (Rappaport, 2011, Wild, 2012). These exposures come from personal care products (Jiménez-Díaz et al., 2014), pharmaceuticals (Bouhifd et al., 2013, Loo et al., 2012), smoking (Jones et al., 2014b, Helen et al., 2012), pollution (Schroijen et al., 2008), plastics (Hogberg et al., 2008, Jonsson et al., 2005) to name a few and these can be detected in tissues and fluids from amniotic fluid and placenta (Vela-Soria et al., 2011, Jensen et al., 2015, Edlow et al., 2012) to urine and blood in adults (Chen et al., 2014, Johnson et al., 2012b, Ye et al., 2011, Schroijen et al., 2008). Due to the vast number and sources of exposure, it is unrealistic to target each chemical contaminant singly, and this makes an omic approach ideal (Wild, 2012). While exposomics is a relatively new field, several metabolomic approaches have been implemented to assess the exposure and its impact upon the organism (Jamin et al., 2014, Al-Salhi et al., 2012). Using a similar SPE methodology to the one developed within this thesis a plasma exposomic approach has been performed on fish plasma and uncovered compounds that were not detectable using previous sample preparations and analytical methods (David et al., 2014). These included low abundance chemical contaminants and their metabolites that were not detected using traditional solvent plasma sample preparation techniques and a conventional UHPLC-ESI-TOF MS analytical platform.

7.6 Concluding remarks

The work presented in this thesis offer a potential solution to increase the sensitivity and detection of low abundance metabolites in the urinary metabolome. These methods were applied to the first urine metabolomic analysis of HIV infection and cART intervention. The use of the newly developed sample preparation techniques and analytical methodologies enabled the detection of novel markers of HIV cART intervention and opened up new areas to investigate and further improve our understanding of HIV pathogenesis and treatment. These studies highlighted the role metabolomics has to offer in medical metabolomic analysis. In a

discovery-based field, the increased use of more sensitive analytical approaches is an exciting prospect as it has the potential to uncover more subtle changes to the metabolome further increasing our understanding of the biochemistry underlying disease or exposure.

References

Aberg, J. A., Tebas, P., Overton, E. T., Gupta, S. K., Sax, P. E., Landay, A., Falcon, R., Ryan, R. & De La Rosa, G. 2012. Metabolic effects of darunavir/ritonavir versus atazanavir/ritonavir in treatment-naive, HIV type 1-infected subjects over 48 weeks. *AIDS Research and Human Retroviruses,* 28**,** 1184-95.

Adamson, D. C., Wildemann, B., Sasaki, M., Glass, J. D., Mcarthur, J. C., Christov, V. I., Dawson, T. M. & Dawson, V. L. 1996. Immunologic NO synthase: elevation in severe AIDS dementia and induction by HIV-1 gp41. *Science,* 274**,** 1917-21.

Agúndez, J. A., García-Martín, E. & Martínez, C. 2009. Genetically based impairment in CYP2C8- and CYP2C9-dependent NSAID metabolism as a risk factor for gastrointestinal bleeding: Is a combination of pharmacogenomics and metabolomics required to improve personalized medicine? *Expert Opinion on Drug Metabolism & Toxicology,* 5**,** 607-620.

Al-Salhi, R., Abdul-Sada, A., Lange, A., Tyler, C. R. & Hill, E. M. 2012. The Xenometabolome and Novel Contaminant Markers in Fish Exposed to a Wastewater Treatment Works Effluent. *Environmental Science & Technology,* 46**,** 9080-9088.

Albert, K. 1995. On-line use of NMR detection in separation chemistry. *Journal of Chromatography A,* 703**,** 123-147.

Aleksunes, L. M., Augustine, L. M., Scheffer, G. L., Cherrington, N. J. & Manautou, J. E. 2008. Renal xenobiotic transporters are differentially expressed in mice following cisplatin treatment. *Toxicology,* 250**,** 82-88.

Alvarez-Sanchez, B., Priego-Capote, F. & De Castro, M. D. L. 2010a. Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. *Trends in Analytical Chemistry,* 29**,** 111-119.

Alvarez-Sanchez, B., Priego-Capote, F. & De Castro, M. D. L. 2010b. Metabolomics analysis II. Preparation of biological samples prior to detection. *Trends in Analytical Chemistry,* 29**,** 120-127.

Amster, I. J. 1996. Fourier transform mass spectrometry. *Journal of Mass Spectrometry,* 31**,** 1325-1337.

Andersen, M. B. S., Reinbach, H. C., Rinnan, A., Barri, T., Mithril, C. & Dragsted, L. O. 2013. Discovery of exposure markers in urine for Brassica-containing meals served with different protein sources by UPLC-qTOF-MS untargeted metabolomics. *Metabolomics,* 9**,** 984-997.

Annaert, P., Ye, Z. W., Stieger, B. & Augustijns, P. 2010. Interaction of HIV protease inhibitors with OATP1B1, 1B3, and 2B1. *Xenobiotica,* 40**,** 163-76.

Annesley, T. M. 2003. Ion suppression in mass spectrometry. *Clinical Chemistry,* 49**,** 1041-1044.

Anumol, T. & Snyder, S. A. 2015. Rapid analysis of trace organic compounds in water by automated online solid-phase extraction coupled to liquid chromatography-tandem mass spectrometry. *Talanta,* 132**,** 77-86.

Arlt, W., Biehl, M., Taylor, A. E., Hahner, S., Libe, R., Hughes, B. A., Schneider, P., Smith, D. J., Stiekema, H., Krone, N., Porfiri, E., Opocher, G., Bertherat, J., Mantero, F., Allolio, B., Terzolo, M., Nightingale, P., Shackleton, C. H. L., Bertagna, X., Fassnacht, M. & Stewart, P. M. 2011. Urine Steroid Metabolomics as a Biomarker Tool for Detecting Malignancy in Adrenal Tumors. *Journal of Clinical Endocrinology & Metabolism,* 96**,** 3775-3784.

-Trevijano, E. R., Lu, S. C., Corrales, F. J. & Mato, J. M. 2004. Methylthioadenosine. *The International Journal of Biochemistry & Cell Biology,* 36**,** 2125-2130.

Balog, J., Sasi-Szabó, L., Kinross, J., Lewis, M. R., Muirhead, L. J., Veselkov, K., Mirnezami, R., Dezső, B., Damjanovich, L., Darzi, A., Nicholson, J. K. & Takáts, Z. 2013. Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. *Science Translational Medicine,* 5**,** 194ra93.

Banday, K. M., Pasikanti, K. K., Chan, E. C. Y., Singla, R., Rao, K. V. S., Chauhan, V. S. & Nanda, R. K. 2011. Use of Urine Volatile Organic Compounds To Discriminate Tuberculosis Patients from Healthy Subjects. *Analytical Chemistry,* 83**,** 5526-5534.

Bathena, S. P. R., Thakare, R., Gautam, N., Mukherjee, S., Olivera, M., Meza, J. & Alnouti, Y. 2015a. Urinary Bile Acids as Biomarkers for Liver Diseases I. Stability of the Baseline Profile in Healthy Subjects. *Toxicological Sciences,* 143**,** 296-307.

Bathena, S. P. R., Thakare, R., Gautam, N., Mukherjee, S., Olivera, M., Meza, J. & Alnouti, Y. 2015b. Urinary Bile Acids as Biomarkers for Liver Diseases II. Signature Profiles in Patients. *Toxicological Sciences,* 143**,** 308-318.

Beger, R. D., Holland, R. D., Sun, J., Schnackenberg, L. K., Moore, P. C., Dent, C. L., Devarajan, P. & Portilla, D. 2008. Metabonomics of acute kidney injury in children after cardiac surgery. *Pediatric Nephrology,* 23**,** 977-984.

Benton, H. P., Want, E., Keun, H. C., Amberg, A., Plumb, R. S., Goldfain-Blanc, F., Walther, B., Reily, M. D., Lindon, J. C., Holmes, E., Nicholson, J. K. & Ebbels, T. M. D. 2012. Intra- and Interlaboratory Reproducibility of Ultra Performance Liquid Chromatography-Time-of-Flight Mass Spectrometry for Urinary Metabolic Profiling. *Analytical Chemistry,* 84**,** 2424-2432.

Berggren, W. T., Westphall, M. S. & Smith, L. M. 2002. Single-pulse nanoelectrospray ionization. *Analytical Chemistry,* 74**,** 3443-8.

Berk, P. D., Howe, R. B., Bloomer, J. R. & Berlin, N. I. 1969. Studies of bilirubin kinetics in normal adults. *Journal of Clinical Investigation,* 48**,** 2176-2190.

Berlioz-Barbier, A., Baudot, R., Wiest, L., Gust, M., Garric, J., Cren-Olivé, C. & Buleté, A. 2015. MicroQuEChERS–nanoliquid chromatography–nanospray–tandem mass spectrometry for the detection and quantification of trace pharmaceuticals in benthic invertebrates. *Talanta,* 132**,** 796-802.

Berlioz-Barbier, A., Buleté, A., Faburé, J., Garric, J., Cren-Olivé, C. & Vulliet, E. 2014. Multi-residue analysis of emerging pollutants in benthic invertebrates by modified micro-quick-easy-cheap-efficient-rugged-safe extraction and nanoliquid chromatography–nanospray–tandem mass spectrometry analysis. *Journal of Chromatography A,* 1367**,** 16-32.

Bingol, K. & Bruschweiler, R. 2014. Multidimensional Approaches to NMR-Based Metabolomics. *Analytical Chemistry,* 86**,** 47-57.

Blask, D. E., Hill, S. M., Dauchy, R. T., Xiang, S. L., Yuan, L., Duplessis, T., Mao, L. L., Dauchy, E. & Sauer, L. A. 2011. Circadian regulation of molecular, dietary, and metabolic signaling mechanisms of human breast cancer growth by the nocturnal melatonin signal and the consequences of its disruption by light at night. *Journal of Pineal Research,* 51**,** 259-269.

Bleakney, W. 1929. A new method of positive ray analysis and its application to the measurement of ionization potentials in mercury vapor. *Physical Review,* 34**,** 157-160.

Boersema, P. J., Mohammed, S. & Heck, A. J. R. 2008. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Analytical and Bioanalytical Chemistry,* 391**,** 151-159.

Bofill, M., Janossy, G., Lee, C. A., Macdonald-Burns, D., Phillips, A. N., Sabin, C., Timms, A., Johnson, M. A. & Kernoff, P. B. A. 1992. Laboratory control values for CD4 and CD8 T lymphocytes. Implications for HIV-1 diagnosis. *Clinical & Experimental Immunology,* 88**,** 243-252.

Boger, M. S., Bian, A., Shintani, A., Milne, G. L., Morrow, J. D., Erdem, H., Mitchell, V., Haas, D. W. & Hulgan, T. 2012. Sex differences in urinary biomarkers of vascular and endothelial function in HIV-infected persons receiving antiretroviral therapy. *Antiviral Therapy,* 17**,** 485-493.

Boisvert, M., Fayad, P. B. & Sauvé, S. 2012. Development of a new multi-residue laser diode thermal desorption atmospheric pressure chemical ionization tandem mass

spectrometry method for the detection and quantification of pesticides and pharmaceuticals in wastewater samples. *Analytica Chimica Acta,* 754**,** 75-82.

Bonvallot, N., Tremblay-Franco, M., Chevrier, C., Canlet, C., Debrauwer, L., Cravedi, J. P. & Cordier, S. 2014. Potential Input From Metabolomics for Exploring and Understanding the Links Between Environment and Health. *Journal of Toxicology and Environmental Health-Part B-Critical Reviews,* 17**,** 21-44.

Bouatra, S., Aziat, F., Mandal, R., Guo, A. C., Wilson, M. R., Knox, C., Bjorndahl, T. C., Krishnamurthy, R., Saleem, F., Liu, P., Dame, Z. T., Poelzer, J., Huynh, J., Yallou, F. S., Psychogios, N., Dong, E., Bogumil, R., Roehring, C. & Wishart, D. S. 2013. The human urine metabolome. *Plos One,* 8**,** e73076.

Boudonck, K. J., Mitchell, M. W., Nemet, L., Keresztes, L., Nyska, A., Shinar, D. & Rosenstock, M. 2009. Discovery of Metabolomics Biomarkers for Early Detection of Nephrotoxicity. *Toxicologic Pathology,* 37**,** 280-292.

Bouhifd, M., Hartung, T., Hogberg, H. T., Kleensang, A. & Zhao, L. 2013. Review: Toxicometabolomics. *Journal of Applied Toxicology,* 33**,** 1365-1383.

Brennan, L. 2014. NMR-based metabolomics: From sample preparation to applications in nutrition research. *Progress in Nuclear Magnetic Resonance Spectroscopy,* 83**,** 42-49.

Broadhurst, D. & Kell, D. 2006. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics,* 2**,** 171-196.

Brown, S. C., Kruppa, G. & Dasseux, J. L. 2005. Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrometry Reviews,* 24**,** 223-231.

Bruggeman, F. J. & Westerhoff, H. V. 2007. The nature of systems biology. *Trends in Microbiology,* 15**,** 45-50.

Bruins, A. P. 1998. Mechanistic aspects of electrospray ionization. *Journal of Chromatography A,* 794**,** 345-357.

Carrola, J., Rocha, C. M., Barros, A. S., Gil, A. M., Goodfellow, B. J., Carreira, I. M., Bernardo, J., Gomes, A., Sousa, V., Carvalho, L. & Duarte, I. F. 2011. Metabolic Signatures of Lung Cancer in Biofluids: NMR-Based Metabonomics of Urine. *Journal of Proteome Research,* 10**,** 221-230.

Cassol, E., Misra, V., Dutta, A., Morgello, S. & Gabuzda, D. 2014. Cerebrospinal fluid metabolomics reveals altered waste clearance and accelerated aging in HIV patients with neurocognitive impairment. *AIDS,* 28**,** 1579-1591.

Cassol, E., Misra, V., Holman, A., Kamat, A., Morgello, S. & Gabuzda, D. 2013. Plasma metabolomics identifies lipid abnormalities linked to markers of inflammation, microbial translocation, and hepatic function in HIV patients receiving protease inhibitors. *BMC Infectious Diseases,* 13.

Cernei, N., Heger, Z., Gumulec, J., Zitka, O., Masarik, M., Babula, P., Eckschlager, T., Stiborova, M., Kizek, R. & Adam, V. 2013. Sarcosine as a Potential Prostate Cancer Biomarker-A Review. *International Journal of Molecular Sciences,* 14**,** 13893-13908.

Chadha, V., Garg, U. & Alon, U. S. 2001. Measurement of urinary concentration: a critical appraisal of methodologies. *Pediatric Nephrology,* 16**,** 374-382.

Chalcraft, K. R. & Mccarry, B. E. 2013. Tandem LC columns for the simultaneous retention of polar and nonpolar molecules in comprehensive metabolomics analysis. *Journal of Separation Science,* 36**,** 3478-3485.

Chandrasekharan, N., Dai, H., Roos, K. L. T., Evanson, N. K., Tomsik, J., Elton, T. S. & Simmons, D. L. 2002. COX-3, a cyclooxygenase-1 variant inhibited by acetaminophen and other analgesic/antipyretic drugs: cloning, structure, and expression. *Proceedings of the National Academy of Sciences,* 99**,** 13926-13931.

Chauve, B., Guillarme, D., Cleon, P. & Veuthey, J. L. 2010. Evaluation of various HILIC materials for the fast separation of polar compounds. *Journal of Separation Science,* 33**,** 752-764.

Chen, M. J., Zhou, K., Chen, X. J., Qiao, S. L., Hu, Y. H., Xu, B., Han, X. M., Tang, R., Mao, Z. L., Dong, C. C., Wu, D., Wang, Y. B., Wang, S. L., Zhou, Z. M., Xia, Y. K. & Wang, X. R. 2014.

Metabolomic Analysis Reveals Metabolic Changes Caused by Bisphenol A in Rats. *Toxicological Sciences,* 138**,** 256-267.

Chen, Y., Xu, J., Zhang, R., Shen, G., Song, Y., Sun, J., He, J., Zhan, Q. & Abliz, Z. 2013a. Assessment of data pre-processing methods for LC-MS/MS-based metabolomics of uterine cervix cancer. *Analyst,* 138**,** 2669-77.

Chen, Y. H., Shen, G. Q., Zhang, R. P., He, J. M., Zhang, Y., Xu, J., Yang, W., Chen, X. G., Song, Y. M. & Abliz, Z. 2013b. Combination of Injection Volume Calibration by Creatinine and MS Signals' Normalization to Overcome Urine Variability in LC-MS-Based Metabolomics Studies. *Analytical Chemistry,* 85**,** 7659-7665.

Cheng, Y., Xie, G., Chen, T., Qiu, Y., Zou, X., Zheng, M., Tan, B., Feng, B., Dong, T., He, P., Zhao, L., Zhao, A., Xu, L. X., Zhang, Y. & Jia, W. 2012. Distinct Urinary Metabolic Profile of Human Colorectal Cancer. *Journal of Proteome Research,* 11**,** 1354-1363.

Chernushevich, I. V., Loboda, A. V. & Thomson, B. A. 2001. An introduction to quadrupole-time-of-flight mass spectrometry. *Journal of Mass Spectrometry,* 36**,** 849-865.

Chertow, G. M., Burdick, E., Honour, M., Bonventre, J. V. & Bates, D. W. 2005. Acute Kidney Injury, Mortality, Length of Stay, and Costs in Hospitalized Patients. *Journal of the American Society of Nephrology,* 16**,** 3365-3370.

Chervet, J. P., Ursem, M. & Salzmann, J. P. 1996. Instrumental requirements for nanoscale liquid chromatography. *Analytical Chemistry,* 68**,** 1507-1512.

Chetwynd, A. J., Abdul-Sada, A. & Hill, E. M. 2015. Solid phase extraction and splitless nanoflow liquid chromatography – nanoelectrospray ionisation mass spectrometry for improved global urine metabolomics. *Analytical Chemistry,* 87**,** 1158-1165.

Chetwynd, A. J., David, A., Abdul-Sada, A. & Hill, E. M. 2014. Evaluation of analytical performance and reliability of direct nanoLC-nanoESI-high resolution mass spectrometry for profiling the (xeno)metabolome. *Journal of Mass Spectrometry,* 49**,** 1063-1069.

Chiang, J. Y. L. 2009. Bile acids: regulation of synthesis. *Journal of Lipid Research,* 50**,** 1955-1966.

Churchwell, M. I., Twaddle, N. C., Meeker, L. R. & Doerge, D. R. 2005. Improving LC-MS sensitivity through increases in chromatographic performance: Comparisons of UPLC-ES/MS/MS to HPLC-ES/MS/MS. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 825**,** 134-143.

Cobb, B. R., Vaks, J. E., Do, T. & Vilchez, R. A. 2011. Evolution in the sensitivity of quantitative HIV-1 viral load tests. *Journal of Clinical Virology,* 52, Supplement 1**,** S77-S82.

Coble, J. B. & Fraga, C. G. 2014. Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery. *Journal of Chromatography A,* 1358**,** 155-164.

Conant, K., Garzino-Demo, A., Nath, A., Mcarthur, J. C., Halliday, W., Power, C., Gallo, R. C. & Major, E. O. 1998. Induction of monocyte chemoattractant protein-1 in HIV-1 Tat-stimulated astrocytes and elevation in AIDS dementia. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 3117-3121.

Connor, S. C., Hansen, M. K., Corner, A., Smith, R. F. & Ryan, T. E. 2010. Integration of metabolomics and transcriptomics data to aid biomarker discovery in type 2 diabetes. *Molecular Biosystems,* 6**,** 909-921.

Creek, D., Dunn, W., Fiehn, O., Griffin, J., Hall, R., Lei, Z., Mistrik, R., Neumann, S., Schymanski, E., Sumner, L., Trengove, R. & Wolfender, J.-L. 2014. Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics,* 10**,** 350-353.

Creek, D. J., Jankevics, A., Burgess, K. E. V., Breitling, R. & Barrett, M. P. 2012. IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data. *Bioinformatics,* 28**,** 1048-1049.

Cribbs, S. K., Park, Y., Guidot, D. M., Martin, G. S., Brown, L. A., Lennox, J. & Jones, D. P. 2014. Metabolomics of bronchoalveolar lavage differentiate healthy HIV-1-infected subjects from controls. *AIDS Research and Human Retroviruses,* 30**,** 579-85.

Crum-Cianflone, N. F., Bavaro, M., Hale, B., Amling, C., Truett, A., Brandt, C., Pope, B., Furtek, K., Medina, S. & Wallace, M. R. 2007. Erectile dysfunction and hypogonadism among men with HIV. *AIDS patient care and STDs,* 21**,** 9-19.

Crum, N. F., Furtek, K. J., Olson, P. E., Amling, C. L. & Wallace, M. R. 2005. A review of hypogonadism and erectile dysfunction among HIV-infected men during the pre-and post-HAART eras: diagnosis, pathogenesis, and management. *AIDS Patient Care & STDs,* 19**,** 655-671.

Cubbon, S., Bradbury, T., Wilson, J. & Thomas-Oates, J. 2007. Hydrophilic interaction chromatography for mass spectrometric metabonomic studies of urine. *Analytical Chemistry,* 79**,** 8911-8918.

Cui, Q., Lewis, I. A., Hegeman, A. D., Anderson, M. E., Li, J., Schulte, C. F., Westler, W. M., Eghbalnia, H. R., Sussman, M. R. & Markley, J. L. 2008. Metabolite identification via the Madison Metabolomics Consortium Database. *Nature Biotechnology,* 26**,** 162-164.

Danielsson, R., Allard, E., Sjoberg, P. J. R. & Bergquist, J. 2011. Exploring liquid chromatography-mass spectrometry fingerprints of urine samples from patients with prostate or urinary bladder cancer. *Chemometrics and Intelligent Laboratory Systems,* 108**,** 33-48.

Dasta, J. F., Kane-Gill, S. L., Durtschi, A. J., Pathak, D. S. & Kellum, J. A. 2008. Costs and outcomes of acute kidney injury (AKI) following cardiac surgery. *Nephrology Dialysis Transplantation,* 23**,** 1970-1974.

Daughton, C. G. & Ruhoy, I. S. 2008. The Afterlife of Drugs and the Role of PharmEcovigilance. *Drug Safety,* 31**,** 1069-1082.

David, A., Abdul-Sada, A., Lange, A., Tyler, C. R. & Hill, E. M. 2014. A new approach for plasma (xeno)metabolomics based on solid-phase extraction and nanoflow liquid chromatography-nanoelectrospray ionisation mass spectrometry. *Journal of Chromatography A,* 24**,** 72-85.

Dawiskiba, T., Deja, S., Mulak, A., Zabek, A., Jawien, E., Pawelka, D., Banasik, M., Mastalerz-Migas, A., Balcerzak, W., Kaliszewski, K., Skora, J., Barc, P., Korta, K., Pormanczuk, K., Szyber, P., Litarski, A. & Mlynarz, P. 2014. Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases. *World Journal of Gastroenterology,* 20**,** 163-174.

De Geus, H. R. H., Betjes, M. G. & Bakker, J. 2012. Biomarkers for the prediction of acute kidney injury: a narrative review on current status and future challenges. *Clinical Kidney Journal,* 5**,** 102-108.

De Hoffmann, E. & Stroobart, V. 2007. *Mass Spectrometry Principals and Applications,* England, John Wiley & Sons Ltd.

Decrion, A. Z., Dichamp, I., Varin, A. & Herbein, G. 2005. HIV and inflammation. *Current HIV Research,* 3**,** 243-259.

Deeks, S. G. 2011. HIV Infection, Inflammation, Immunosenescence, and Aging. *Annual Review of Medicine,* 62**,** 141-155.

Deeks, S. G. & Phillips, A. N. 2009. HIV infection, antiretroviral treatment, ageing, and non-AIDS related morbidity. *British Medical Journal,* 338.

Delanghe, J. & Speeckaert, M. 2014. Preanalytical requirements of urinalysis. *Biochemia Medica,* 24**,** 89-104.

Dettmer, K., Aronov, P. A. & Hammock, B. D. 2007. Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews,* 26**,** 51-78.

Ding, J. M. & Anderegg, R. J. 1995. Specific and nonspecific dimer formation in the electrospray ionization mass spectrometry of oligonucleotides. *Journal of the American Society for Mass Spectrometry,* 6**,** 159-164.

Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R. & Griffin, J. L. 2011. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews,* 40**,** 387-426.

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J. & Viant, M. R. 2013. Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics,* 9**,** 44-66.

Edlow, A. G., Chen, M., Smith, N. A., Lu, C. & Mcelrath, T. F. 2012. Fetal bisphenol A exposure: Concentration of conjugated and unconjugated bisphenol A in amniotic fluid in the second and third trimesters. *Reproductive Toxicology,* 34**,** 1-7.

Edmands, W. M. B., Beckonert, O. P., Stella, C., Campbell, A., Lake, B. G., Lindon, J. C., Holmes, E. & Gooderham, N. J. 2011. Identification of Human Urinary Biomarkers of Cruciferous Vegetable Consumption by Metabonomic Profiling. *Journal of Proteome Research,* 10**,** 4513-4521.

Edmands, W. M. B., Ferrari, P. & Scalbert, A. 2014. Normalization to specific gravity prior to analysis improves information recovery from high resolution mass spectrometry metabolomic profiles of human urine. *Analytical Chemistry*.

Edwards, J. L., Edwards, R. L., Reid, K. R. & Kennedy, R. T. 2007. Effect of decreasing column inner diameter and use of off-line two-dimensional chromatography on metabolite detection in complex mixtures. *Journal of Chromatography A,* 23**,** 127-34.

Ek, P., Stjernstrom, M., Emmer, A. & Roeraade, J. 2010. Electrospray ionization mass spectrometry from discrete nanoliter-sized sample volumes. *Rapid Communications in Mass Spectrometry,* 24**,** 2561-2568.

El Nahas, A. M. & Bello, A. K. 2005. Chronic kidney disease: the global challenge. *The Lancet,* 365**,** 331-340.

Emmett, M. R. & Caprioli, R. M. 1994. Micro-Electrospray Mass-Spectrometry - Ultra-High-Sensitivity Analysis Of Peptides And Proteins. *Journal of the American Society for Mass Spectrometry,* 5**,** 605-613.

Emwas, A. H. M., Salek, R. M., Griffin, J. L. & Merzaban, J. 2013. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. *Metabolomics,* 9**,** 1048-1072.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikstrom, C. & S., W. 2006. *Multi- and Megavariate Data Analysis Part 1 Basic Principals and Applications*, Umetrics AB.

Exarchou, V., Krucker, M., Van Beek, T. A., Vervoort, J., Gerothanassis, I. P. & Albert, K. 2005. LC–NMR coupling technology: recent advancements and applications in natural products analysis. *Magnetic Resonance in Chemistry,* 43**,** 681-687.

Fanali, S., Aturki, Z., D'orazio, G. & Rocco, A. 2007. Separation of basic compounds of pharmaceutical interest by using nano-liquid chromatography coupled with mass spectrometry. *Journal of Chromatography A,* 1150**,** 252-8.

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O. G. & Lemey, P. 2014. The early spread and epidemic ignition of HIV-1 in human populations. *Science,* 346**,** 56-61.

Fauci, A. S. 1996. Host factors and the pathogenesis of HIV-induced disease. *Nature,* 384**,** 529-534.

Fave, G., Beckmann, M. E., Draper, J. H. & Mathers, J. C. 2009. Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? *Genes and Nutrition,* 4**,** 135-141.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 246**,** 64-71.

Fenn, L., Kliman, M., Mahsut, A., Zhao, S. & Mclean, J. 2009. Characterizing ion mobility-mass spectrometry conformation space for the analysis of complex biological samples. *Analytical and Bioanalytical Chemistry,* 394**,** 235-244.

Fernández-Arroyo, S., Gómez-Martínez, A., Rocamora-Reverte, L., Quirantes-Piné, R., Segura-Carretero, A., Fernández-Gutiérrez, A. & Ferragut, J. A. 2012. Application of nanoLC-ESI-TOF-MS for the metabolomic analysis of phenolic compounds from extra-virgin olive oil in treated colon-cancer cells. *Journal of Pharmaceutical and Biomedical Analysis,* 63**,** 128-134.

Fernández-Peralbo, M. A. & Luque De Castro, M. D. 2012. Preparation of urine samples prior to targeted or untargeted metabolomics mass-spectrometry analysis. *Trends in Analytical Chemistry,* 41**,** 75-85.

Fiehn, O. 2002. Metabolomics - the link between genotypes and phenotypes. *Plant Molecular Biology,* 48**,** 155-171.

Fischer, K., Kettunen, J., Würtz, P., Haller, T., Havulinna, A., Kangas, A., Soininen, P., Esko, T., Tammesoo, M.-L., Mägi, R., Smit, S., Palotie, A., Ripatti, S., Salomaa, V., Ala-Korpela, M., Perola, M. & Metspalu, A. 2014. Biomarker Profiling by Nuclear Magnetic Resonance Spectroscopy for the Prediction of All-Cause Mortality: An Observational Study of 17,345 Persons. *PLOS Medicine,* 11**,** e1001606.

Fischer, R., Bowness, P. & Kessler, B. M. 2013. Two birds with one stone: Doing metabolomics with your proteomics kit. *Proteomics,* 13**,** 3371-3386.

Fitzgerald, D. W., Bezak, K., Ocheretina, O., Riviere, C., Wright, T. C., Milne, G. L., Zhou, X. K., Du, B., Subbaramaiah, K., Byrt, E., Goodwin, M. L., Rafii, A. & Dannenberg, A. J. 2012. The Effect of HIV and HPV Coinfection on Cervical COX-2 Expression and Systemic Prostaglandin E2 Levels. *Cancer Prevention Research,* 5**,** 34-40.

Foley, P., Kazazi, F., Biti, R., Sorrell, T. C. & Cunningham, A. L. 1992. HIV infection of monocytes inhibits the T-lymphocyte proliferative response to recall antigens, via production of eicosanoids. *Immunology,* 75**,** 391-397.

Froldi, M., Castagna, A., Parma, M., Piona, A., Tedeschi, A., Miadonna, A., Lorini, M., Orazio, E. N. & Lazzarin, A. 1992. Mediator release in cerebrospinal fluid of human immunodeficiency virus-positive patients with central nervous system involvement. *Journal of Neuroimmunology,* 38**,** 155-61.

Fuhrer, T. & Zamboni, N. 2015. High-throughput discovery metabolomics. *Current Opinion in Biotechnology,* 31**,** 73-78.

Gallant, J. E. & Deresinski, S. 2003. Tenofovir Disoproxil Fumarate. *Clinical Infectious Diseases,* 37**,** 944-950.

Gama, M. R., Collins, C. H. & Bottoli, C. B. G. 2013. Nano-Liquid Chromatography in Pharmaceutical and Biomedical Research. *Journal of Chromatographic Science,* 51**,** 694-703.

Gangl, E. T., Annan, M. M., Spooner, N. & Vouros, P. 2001. Reduction of signal suppression effects in ESI-MS using a nanosplitting device. *Analytical Chemistry,* 73**,** 5635-44.

Ganti, S., Taylor, S. L., Abu Aboud, O., Yang, J., Evans, C., Osier, M. V., Alexander, D. C., Kim, K. & Weiss, R. H. 2012a. Kidney tumor biomarkers revealed by simultaneous multiple matrix metabolomics analysis. *Cancer Research,* 72**,** 3471-9.

Ganti, S., Taylor, S. L., Kim, K., Hoppel, C. L., Guo, L., Yang, J., Evans, C. & Weiss, R. H. 2012b. Urinary acylcarnitines are altered in human kidney cancer. *International Journal of Cancer,* 130**,** 2791-2800.

Ganti, S. & Weiss, R. H. 2011. Urine metabolomics for kidney cancer detection and biomarker discovery. *Urologic Oncology-Seminars and Original Investigations,* 29**,** 551-557.

Gebregiworgis, T., Massilamany, C., Gangaplara, A., Thulasingam, S., Kolli, V., Werth, M. T., Dodds, E. D., Steffen, D., Reddy, J. & Powers, R. 2013. Potential of Urinary Metabolites for Diagnosing Multiple Sclerosis. *ACS Chemical Biology,* 8**,** 684-690.

Genta-Jouve, G., Croué, J., Weinberg, L., Cocandeau, V., Holderith, S., Bontemps, N., Suzuki, M. & Thomas, O. P. 2014. Two-dimensional ultra high pressure liquid chromatography quadrupole/time-of-flight mass spectrometry for semi-targeted natural compounds identification. *Phytochemistry Letters,* 10**,** 318-323.

German, J. B., Hammock, B. D. & Watkins, S. M. 2005. Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics,* 1**,** 3-9.

Ghannoum, M. A., Mukherjee, P. K., Jurevic, R. J., Retuerto, M., Brown, R. E., Sikaroodi, M., Webster-Cyriaque, J. & Gillevet, P. M. 2013. Metabolomics Reveals Differential Levels of Oral Metabolites in HIV-Infected Patients: Toward Novel Diagnostic Targets. *Omics-a Journal of Integrative Biology,* 17**,** 5-15.

Gika, H. G., Theodoridis, G. A., Plumb, R. S. & Wilson, I. D. 2014. Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *Journal of Pharmaceutical and Biomedical Analysis,* 87**,** 12-25.

Gika, H. G., Theodoridis, G. A. & Wilson, I. D. 2008a. Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography TOF-MS for metabonomic analysis of Zucker rat urine. *Journal of Separation Science,* 31**,** 1598-1608.

Gika, H. G., Theodoridis, G. A. & Wilson, I. D. 2008b. Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human urine - Sample stability under different handling and storage conditions for metabonomics studies. *Journal of Chromatography A,* 1189**,** 314-322.

Gika, H. G., Theodoridis, G. A., Wingate, J. E. & Wilson, I. D. 2007. Within-day reproducibility of an HPLC-MS-Based method for metabonomic analysis: Application to human urine. *Journal of Proteome Research,* 6**,** 3291-3303.

Glish, G. L. & Vachet, R. W. 2003. The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery,* 2**,** 140-150.

Godzien, J., García-Martínez, D., Martinez-Alcazar, P., Ruperez, F. & Barbas, C. 2013. Effect of a nutraceutical treatment on diabetic rats with targeted and CE-MS non-targeted approaches. *Metabolomics,* 9**,** 188-202.

Gomez, C. & Hope, T. J. 2005. The ins and outs of HIV replication. *Cell Microbiol,* 7**,** 621-6.

Goodacre, R. 2007. Metabolomics of a superorganism. *Journal of Nutrition,* 137**,** 259S-266S.

Gorshkov, M. V. & Zubarev, R. A. 2005. *On the accuracy of polypeptide masses measured in a linear ion trap*, Rapid Communications Mass Spectrometry. 2005;19(24):3755-8.

Gouveia, M. J., Brindley, P. J., Santos, L. L., Da Costa, J. M. C., Gomes, P. & Vale, N. 2013. Mass spectrometry techniques in the survey of steroid metabolites as potential disease biomarkers: A review. *Metabolism-Clinical and Experimental,* 62**,** 1206-1217.

Greene, W. C. 1991. MECHANISMS OF DISEASE - THE MOLECULAR-BIOLOGY OF HUMAN-IMMUNODEFICIENCY-VIRUS TYPE-1 INFECTION. *New England Journal of Medicine,* 324**,** 308-317.

Griffin, D. E., Wesselingh, S. L. & Mcarthur, J. C. 1994. Elevated central nervous system prostaglandins in human immunodeficiency virus-associated dementia. *Annals of Neurology,* 35**,** 592-7.

Griffin, L., Annaert, P. & Brouwer, K. 2011. Influence of Drug Transport Proteins on Pharmacokinetics and Drug Interactions of HIV Protease Inhibitors. *Journal of Pharmaceutical Sciences,* 100**,** 3636-3654.

Griffin, L. M., Watkins, P. B., Perry, C. H., St. Claire, R. L. & Brouwer, K. L. R. 2013. Combination Lopinavir and Ritonavir Alter Exogenous and Endogenous Bile Acid Disposition in Sandwich-Cultured Rat Hepatocytes. *Drug Metabolism and Disposition,* 41**,** 188-196.

Griffiths, W. J., Karua, K., Hornshaw, M., Woffendin, G. & Wanga, Y. 2007. Metabolomics and metabolite profiling: past heroes and future developments. *European Journal of Mass Spectrometry,* 13**,** 45-50.

Gu, H. W., Chen, H. W., Pan, Z. Z., Jackson, A. U., Talaty, N., Xi, B. W., Kissinger, C., Duda, C., Mann, D., Raftery, D. & Cooks, R. G. 2007. Monitoring diet effects via biofluids and their implications for metabolomics studies. *Analytical Chemistry,* 79**,** 89-97.

Guilhaus, M. 1995. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. *Journal of Mass Spectrometry,* 30**,** 1519-1532.

Guo, K., Peng, J., Zhou, R. & Li, L. 2011. Ion-pairing reversed-phase liquid chromatography fractionation in combination with isotope labeling reversed-phase liquid chromatography–mass spectrometry for comprehensive metabolome profiling. *Journal of Chromatography A,* 1218**,** 3689-3694.

Guy, P. A., Tavazzi, I., Bruce, S. J., Ramadan, Z. & Kochhar, S. 2008. Global metabolic profiling analysis on human urine by UPLC-TOFMS: Issues and method validation in nutritional metabolomics. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 871**,** 253-260.

Hagan, S. O., Dunn, W. B., Knowles, J. D., Broadhurst, D., Williams, R., Ashworth, J. J., Cameron, M. & Kell, D. B. 2007. Closed-loop, multiobjective optimization of two-dimensional gas chromatography/mass spectrometry for serum metabolomics. *Analytical Chemistry,* 79**,** 464-476.

Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. 2000. AIDS - AIDS as a zoonosis: Scientific and public health implications. *Science,* 287**,** 607-614.

Hall, A. M., Hendry, B. M., Nitsch, D. & Connolly, J. O. 2011a. Tenofovir-Associated Kidney Toxicity in HIV-Infected Patients: A Review of the Evidence. *American Journal of Kidney Diseases,* 57**,** 773-780.

Hall, J. A., Brockman, J. A. & Jewell, D. E. 2011b. Dietary fish oil alters the lysophospholipid metabolomic profile and decreases urinary 11-dehydro thromboxane B-2 concentration in healthy Beagles. *Veterinary Immunology and Immunopathology,* 144**,** 355-365.

Harrison, A. G. 1992. *Chemical ionization mass spectrometry*, CRC press.

Haseltine, W. A. 1991. Molecular biology of the human immunodeficiency virus type 1. *Faseb Journal,* 5**,** 2349-2360.

Hawkey, C. J. 1999. COX-2 inhibitors. *The Lancet,* 353**,** 307-314.

Hayashi, K., Sasamura, H., Hishiki, T., Suematsu, M., Ikeda, S., Soga, T. & Itoh, H. 2011. Use of serum and urine metabolome analysis for the detection of metabolic changes in patients with stage 1-2 chronic kidney disease. *Nephro-Urology Monthly,* 03**,** 164-171.

Heavner, D. L., Morgan, W. T., Sears, S. B., Richardson, J. D., Byrd, G. D. & Ogden, M. W. 2006. Effect of creatinine and specific gravity normalization techniques on xenobiotic biomarkers in smokers' spot and 24-h urines. *Journal of Pharmaceutical and Biomedical Analysis,* 40**,** 928-942.

Hegde, P. S., White, I. R. & Debouck, C. 2003. Interplay of transcriptomics and proteomics. *Current Opinion in Biotechnology,* 14**,** 647-651.

Heinegård, D. & Tiderström, G. 1973. Determination of serum creatinine by a direct colorimetric method. *Clinica Chimica Acta,* 43**,** 305-310.

Helen, G., Bernert, J. T., Hall, D. B., Sosnoff, C. S., Xia, Y., Balmes, J. R., Vena, J. E., Wang, J. S., Holland, N. T. & Naeher, L. P. 2012. Exposure to Secondhand Smoke Outside of a Bar and a Restaurant and Tobacco Exposure Biomarkers in Nonsmokers. *Environmental Health Perspectives,* 120**,** 1010-1016.

Hemstrom, P. & Irgum, K. 2006. Hydrophilic interaction chromatography. *Journal of Separation Science,* 29**,** 1784-1821.

Hernandez-Borges, J., Aturki, Z., Rocco, A. & Fanali, S. 2007. Recent applications in nanoliquid chromatography. *Journal of Separation Science,* 30**,** 1589-1610.

Hill, A., Van Der Lugt, J., Sawyer, W. & Boffito, M. 2009. How much ritonavir is needed to boost protease inhibitors? Systematic review of 17 dose-ranging pharmacokinetic trials. *AIDS,* 23**,** 2237-2245

Hogberg, J., Hanberg, A., Berglund, M., Skerfving, S., Remberger, M., Calafat, A. M., Filipsson, A. F., Jansson, B., Johansson, N., Appelgren, M. & Hakansson, H. 2008. Phthalate diesters and their metabolites in human breast milk, blood or serum, and urine as biomarkers of exposure in vulnerable populations. *Environmental Health Perspectives,* 116**,** 334-339.

Hollenbaugh, J. A., Munger, J. & Kim, B. 2011. Metabolite profiles of human immunodeficiency virus infected CD4+T cells and macrophages using LC-MS/MS analysis. *Virology,* 415**,** 153-159.

Holmes, E., Loo, R. L., Cloarec, O., Coen, M., Tang, H. R., Maibaum, E., Bruce, S., Chan, Q., Elliott, P., Stamler, J., Wilson, I. D., Lindon, J. C. & Nicholson, J. K. 2007. Detection of urinary drug metabolite (Xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (NMR) spectroscopy. *Analytical Chemistry,* 79**,** 2629-2640.

Holmes, E., Loo, R. L., Stamler, J., Bictash, M., Yap, I. K. S., Chan, Q., Ebbels, T., De Iorio, M., Brown, I. J., Veselkov, K. A., Daviglus, M. L., Kesteloot, H., Ueshima, H., Zhao, L. C., Nicholson, J. K. & Elliott, P. 2008. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature,* 453**,** 396-U50.

Hommes, M. J., Romijn, J. A., Endert, E. & Sauerwein, H. P. 1991. Resting energy expenditure and substrate oxidation in human immunodeficiency virus (HIV)-infected asymptomatic men: HIV affects host metabolism in the early asymptomatic stage. *American Journal of Clinical Nutrition,* 54**,** 311-5.

Hommes, M. J., Romijn, J. A., Godfried, M. H., Schattenkerk, J. K., Buurman, W. A., Endert, E. & Sauerwein, H. P. 1990. Increased resting energy expenditure in human immunodeficiency virus-infected men. *Metabolism-Clinical and Experimental,* 39**,** 1186-90.

Hsu, A., Granneman, G. R., Witt, G., Locke, C., Denissen, J., Molla, A., Valdes, J., Smith, J., Erdman, K. & Lyons, N. 1997. Multiple-dose pharmacokinetics of ritonavir in human immunodeficiency virus-infected subjects. *Antimicrobial agents and chemotherapy,* 41**,** 898-905.

Hu, Q. Z., Noll, R. J., Li, H. Y., Makarov, A., Hardman, M. & Cooks, R. G. 2005. The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry,* 40**,** 430-443.

Huang, Z., Lin, L., Gao, Y., Chen, Y., Yan, X., Xing, J. & Hang, W. 2011. Bladder Cancer Determination Via Two Urinary Metabolites: A Biomarker Pattern Approach. *Molecular & Cellular Proteomics,* 10.

Huang, Z. Z., Chen, Y. J., Hang, W., Gao, Y., Lin, L., Li, D. Y., Xing, J. C. & Yan, X. M. 2013. Holistic metabonomic profiling of urine affords potential early diagnosis for bladder and kidney cancers. *Metabolomics,* 9**,** 119-129.

Hulgan, T., Morrow, J., D'aquila, R. T., Raffanti, S., Morgan, M., Rebeiro, P. & Haas, D. W. 2003. Oxidant Stress Is Increased during Treatment of Human Immunodeficiency Virus Infection. *Clinical Infectious Diseases,* 37**,** 1711-1717.

Huls, M., Van Den Heuvel, J. J., Dijkman, H. B., Russel, F. G. & Masereeuw, R. 2006. ABC transporter expression profiling after ischemic reperfusion injury in mouse kidney. *Kidney International,* 69**,** 2186-93.

Hunt, P. 2012. HIV and Inflammation: Mechanisms and Consequences. *Current HIV/AIDS Reports,* 9**,** 139-147.

Husdan, H. & Rapoport, A. 1968. Estimation of creatinine by the Jaffe reaction. A comparison of three methods. *Clinical Chemistry,* 14**,** 222-38.

Ibrahim, Y. M., Baker, E. S., Danielson Iii, W. F., Norheim, R. V., Prior, D. C., Anderson, G. A., Belov, M. E. & Smith, R. D. Development of a new ion mobility time-of-flight mass spectrometer. *International Journal of Mass Spectrometry*.

Idborg, H., Zamani, L., Edlund, P. O., Schuppe-Koistinen, I. & Jacobsson, S. P. 2005. Metabolic fingerprinting of rat urine by LC/MS Part 1. Analysis by hydrophilic interaction liquid

chromatography-electrospray ionization mass spectrometry. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 828**,** 9-13.

Issaq, H. J., Abbott, E. & Veenstra, T. D. 2008a. Utility of separation science in metabolomic studies. *Journal of Separation Science,* 31**,** 1936-1947.

Issaq, H. J., Nativ, O., Waybright, T., Luke, B., Veenstra, T. D., Issaq, E. J., Kravstov, A. & Mullerad, M. 2008b. Detection of bladder cancer in human urine by metabolomic profiling using high performance liquid chromatography/mass spectrometry. *Journal of Urology,* 179**,** 2422-2426.

Issaq, H. J., Van, Q. N., Waybright, T. J., Muschik, G. M. & Veenstra, T. D. 2009. Analytical and statistical approaches to metabolomics research. *Journal of Separation Science,* 32**,** 2183-2199.

Jacob, C., Dervilly-Pinel, G., Biancotto, G., Monteau, F. & Le Bizec, B. 2014a. Global urine fingerprinting by LC-ESI(+)-HRMS for better characterization of metabolic pathway disruption upon anabolic practices in bovine. *Metabolomics***,** 1-14.

Jacob, C., Dervilly-Pinel, G., Biancotto, G., Monteau, F. & Le Bizec, B. 2015. Global urine fingerprinting by LC-ESI(+)-HRMS for better characterization of metabolic pathway disruption upon anabolic practices in bovine. *Metabolomics,* 11**,** 184-197.

Jacob, C. C., Dervilly-Pinel, G., Biancotto, G. & Le Bizec, B. 2014b. Evaluation of specific gravity as normalization strategy for cattle urinary metabolome analysis. *Metabolomics,* 10**,** 627-637.

Jacobs, D. M., Fuhrmann, J. C., Van Dorsten, F. A., Rein, D., Peters, S., Van Velzen, E. J. J., Hollebrands, B., Draijer, R., Van Duynhoven, J. & Garczarek, U. 2012. Impact of Short-Term Intake of Red Wine and Grape Polyphenol Extract on the Human Metabolome. *Journal of Agricultural and Food Chemistry,* 60**,** 3078-3085.

Jain, R. G., Furfine, E. S., Pedneault, L., White, A. J. & Lenhard, J. M. 2001. Metabolic complications associated with antiretroviral therapy. *Antiviral Research,* 51**,** 151-177.

Jamin, E. L., Bonvallot, N., Tremblay-Franco, M., Cravedi, J. P., Chevrier, C., Cordier, S. & Debrauwer, L. 2014. Untargeted profiling of pesticide metabolites by LC-HRMS: an exposomics tool for human exposure evaluation. *Analytical and Bioanalytical Chemistry,* 406**,** 1149-1161.

Jandera, P. 2012. Comprehensive two-dimensional liquid chromatography — practical impacts of theoretical considerations. A review. *Central European Journal of Chemistry,* 10**,** 844-875.

Jensen, M. S., Anand-Ivell, R., Nørgaard-Pedersen, B., Jönsson, B. a. G., Bonde, J. P., Hougaard, D. M., Cohen, A., Lindh, C. H., Ivell, R. & Toft, G. 2015. Amniotic Fluid Phthalate Levels and Male Fetal Gonad Function. *Epidemiology,* 26**,** 91-99.

Jiménez-Díaz, I., Zafra-Gómez, A., Ballesteros, O. & Navalón, A. 2014. Analytical methods for the determination of personal care products in human samples: An overview. *Talanta,* 129**,** 448-458.

Johnson, C. H., Patterson, A. D., Idle, J. R. & Gonzalez, F. J. 2012a. Xenobiotic Metabolomics: Major Impact on the Metabolome. *Annual Review of Pharmacology and Toxicology, Vol 52,* 52**,** 37-56.

Johnson, C. H., Patterson, A. D., Krausz, K. W., Kalinich, J. F., Tyburski, J. B., Kang, D. W., Luecke, H., Gonzalez, F. J., Blakely, W. F. & Idle, J. R. 2012b. Radiation Metabolomics. 5. Identification of Urinary Biomarkers of Ionizing Radiation Exposure in Nonhuman Primates by Mass Spectrometry-Based Metabolomics. *Radiation Research,* 178**,** 328-340.

Jones, D. R., Wu, Z. P., Chauhan, D., Anderson, K. C. & Peng, J. M. 2014a. A Nano Ultra-Performance Liquid Chromatography-High Resolution Mass Spectrometry Approach for Global Metabolomic Profiling and Case Study on Drug-Resistant Multiple Myeloma. *Analytical Chemistry,* 86**,** 3667-3675.

Jones, I. A., St Helen, G., Meyers, M. J., Dempsey, D. A., Havel, C., Jacob, P., Northcross, A., Hammond, S. K. & Benowitz, N. L. 2014b. Biomarkers of secondhand smoke exposure in automobiles. *Tobacco Control,* 23**,** 51-57.

Jonsson, B. a. G., Richthoff, J., Rylander, L., Giwercman, A. & Hagmar, L. 2005. Urinary phthalate metabolites and biomarkers of reproductive function in young men. *Epidemiology,* 16**,** 487-493.

Juraschek, R., Dulcks, T. & Karas, M. 1999. Nanoelectrospray--more than just a minimized-flow electrospray ionization source. *Journal of the American Society for Mass Spectrometry,* 10**,** 300-8.

Kaddurah-Daouk, R., Kristal, B. S. & Weinshilboum, R. M. 2008. Metabolomics: A global biochemical approach to drug response and disease. *Annual Review of Pharmacology and Toxicology.* Palo Alto: Annual Reviews.

Kalliokoski, A. & Niemi, M. 2009. Impact of OATP transporters on pharmacokinetics. *British Journal of Pharmacology,* 158**,** 693-705.

Kamat, A., Misra, V., Cassol, E., Ancuta, P., Yan, Z., Li, C., Morgello, S. & Gabuzda, D. 2012. A plasma biomarker signature of immune activation in HIV patients on antiretroviral therapy. *Plos One,* 7**,** 17.

Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. & Hill, H. H. 2008. Ion mobility–mass spectrometry. *Journal of Mass Spectrometry,* 43**,** 1-22.

Kanu, A. B., Kumar, B. S. & Hill, H. H. 2012. Evaluation of micro-versus nano-electrospray ionization for ambient pressure ion mobility spectrometry. *International Journal for Ion Mobility Spectrometry,* 15**,** 9-20.

Karas, M., Bahr, U. & Dulcks, T. 2000. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius Journal of Analytical Chemistry,* 366**,** 669-676.

Katajamaa, M. & Oresic, M. 2005. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics,* 6**,** 179.

Katajamaa, M. & Orešič, M. 2007. Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A,* 1158**,** 318-328.

Kearney, B., Flaherty, J. & Shah, J. 2004. Tenofovir Disoproxil Fumarate. *Clinical Pharmacokinetics,* 43**,** 595-612.

Kell, D. B. 2006. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discovery Today,* 11**,** 1085-1092.

Kell, D. B. & Goodacre, R. 2014. Metabolomics and systems pharmacology: why and how to model the human metabolic network for drug discovery. *Drug Discovery Today,* 19**,** 171-182.

Kim, K., Taylor, S. L., Ganti, S., Guo, L., Osier, M. V. & Weiss, R. H. 2011. Urine Metabolomic Analysis Identifies Potential Biomarkers and Pathogenic Pathways in Kidney Cancer. *Omics-a Journal of Integrative Biology,* 15**,** 293-303.

Kind, T., Tolstikov, V., Fiehn, O. & Weiss, R. H. 2007. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Analytical Biochemistry,* 363**,** 185-195.

Kohler, J. J., Hosseini, S. H., Hoying-Brandt, A., Green, E., Johnson, D. M., Russ, R., Tran, D., Raper, C. M., Santoianni, R. & Lewis, W. 2009. Tenofovir renal toxicity targets mitochondria of renal proximal tubules. *Laboratory Investigation,* 89**,** 513-519.

Koppen, V., Jones, R., Bockx, M. & Cuyckens, F. 2014. High volume injections of biological samples for sensitive metabolite profiling and quantitation. *Journal of Chromatography A,* 1372**,** 102-109.

Koppenaal, D. W., Barinaga, C. J., Denton, M. B., Sperline, R. P., Hieftje, G. M., Schilling, G. D., Andrade, F. J. & Barnes, J. H. 2005. Ms detectors. *Analytical Chemistry,* 77**,** 418A-427A.

Kosmiski, L. 2011. Energy expenditure in HIV infection. *American Journal of Clinical Nutrition,* 94**,** 16.

Kuharev, J., Navarro, P., Distler, U., Jahn, O. & Tenzer, S. 2015. In-depth evaluation of software tools for data-independent acquisition based label-free quantification. *Proteomics***,** doi: 10.1002/pmic.201400396.

Kumar, B. S., Chung, B. C., Kwon, O.-S. & Jung, B. H. 2012. Discovery of common urinary biomarkers for hepatotoxicity induced by carbon tetrachloride, acetaminophen and methotrexate by mass spectrometry-based metabolomics. *Journal of Applied Toxicology,* 32**,** 505-520.

Kussmann, M., Raymond, F. & Affolter, M. 2006. OMICS-driven biomarker discovery in nutrition and health. *Journal of Biotechnology,* 124**,** 758-787.

Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. 2014. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature Chemistry,* 6**,** 281-294.

Le Guennec, A., Giraudeau, P. & Caldarelli, S. 2014. Evaluation of Fast 2D NMR for Metabolomics. *Analytical Chemistry,* 86**,** 5946-5954.

Le Tiec, C., Barrail, A., Goujard, C. & Taburet, A.-M. 2005. Clinical pharmacokinetics and summary of efficacy and tolerability of atazanavir. *Clinical pharmacokinetics,* 44**,** 1035-1050.

Lee, S. H. & Cho, Y. D. 1998. Induction of Apoptosis in Leukemia U937 Cells by 5′-Deoxy-5′-methylthioadenosine, a Potent Inhibitor of Protein Carboxylmethyltransferase. *Experimental Cell Research,* 240**,** 282-292.

Lefebvre, P., Cariou, B., Lien, F., Kuipers, F. & Staels, B. 2009. *Role of Bile Acids and Bile Acid Receptors in Metabolic Regulation*.

Lenz, E. M., Bright, J., Wilson, I. D., Hughes, A., Morrisson, J., Lindberg, H. & Lockton, A. 2004. Metabonomics, dietary influences and cultural differences: a H-1 NMR-based study of urine samples obtained from healthy British and Swedish subjects. *Journal of Pharmaceutical and Biomedical Analysis,* 36**,** 841-849.

Lenz, E. M., Bright, J., Wilson, I. D., Morgan, S. R. & Nash, A. F. P. 2003. A H-1 NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. *Journal of Pharmaceutical and Biomedical Analysis,* 33**,** 1103-1115.

Lenz, E. M. & Wilson, I. D. 2007. Analytical strategies in metabonomics. *Journal of Proteome Research,* 6**,** 443-458.

Levey, A. S., Coresh, J., Balk, E., Kausz, A. T., Levin, A., Steffes, M. W., Hogg, R. J., Perrone, R. D., Lau, J. & Eknoyan, G. 2003. National kidney foundation practice guidelines for chronic kidney disease: Evaluation, classification, and stratification. *Annals of Internal Medicine,* 139**,** 137-147.

Levy, J. A. 1993. Pathogenesis of Human-Immunodeficiency-Virus Infection. *Microbiological Reviews,* 57**,** 183-289.

Li, L., Li, R., Zhou, J., Zuniga, A., Stanislaus, A. E., Wu, Y., Huan, T., Zheng, J., Shi, Y., Wishart, D. S. & Lin, G. 2013a. MyCompoundID: Using an Evidence-based Metabolome Library for Metabolite Identification. *Analytical Chemistry,* 85**,** 3401-3408.

Li, M., Feng, B., Liang, Y., Zhang, W., Bai, Y., Tang, W., Wang, T. & Liu, H. 2013b. Lipid profiling of human plasma from peritoneal dialysis patients using an improved 2D (NP/RP) LC-QToF MS method. *Analytical and Bioanalytical Chemistry,* 405**,** 6629-6638.

Lien, G. W., Chen, C. Y. & Wang, G. S. 2009. Comparison of electrospray ionization, atmospheric pressure chemical ionization and atmospheric pressure photoionization for determining estrogenic chemicals in water by liquid chromatography tandem mass spectrometry with chemical derivatizations. *Journal of Chromatography A,* 1216**,** 956-966.

Liland, K. H. 2011. Multivariate methods in metabolomics - from pre-processing to dimension reduction and statistical analysis. *Trac-Trends in Analytical Chemistry,* 30**,** 827-841.

Lin, H.-L., D'agostino, J., Kenaan, C., Calinski, D. & Hollenberg, P. F. 2013. The Effect of Ritonavir on Human CYP2B6 Catalytic Activity: Heme Modification Contributes to the

Mechanism-Based Inactivation of CYP2B6 and CYP3A4 by Ritonavir. *Drug Metabolism and Disposition,* 41**,** 1813-1824.

Lin, H.-M., Edmunds, S. J., Helsby, N. A., Ferguson, L. R. & Rowan, D. D. 2009. Nontargeted Urinary Metabolite Profiling of a Mouse Model of Crohn's Disease. *Journal of Proteome Research,* 8**,** 2045-2057.

Lindon, J. C., Holmes, E. & Nicholson, J. K. 2007. Metabonomics in pharmaceutical R & D. *FEBS Journal,* 274**,** 1140-1151.

Liu, J. Z., Zhao, Z. Y. & Teffera, Y. 2012. Application of on-line nano-liquid chromatography/mass spectrometry in metabolite identification studies. *Rapid Communications in Mass Spectrometry,* 26**,** 320-326.

Liu, L., Aa, J., Wang, G., Yan, B., Zhang, Y., Wang, X., Zhao, C., Cao, B., Shi, J., Li, M., Zheng, T., Zheng, Y., Hao, G., Zhou, F., Sun, J. & Wu, Z. 2010. Differences in metabolite profile between blood plasma and serum. *Analytical Biochemistry,* 406**,** 105-12.

Liu, S., Sun, M. Z., Tang, J. W., Wang, Z., Sun, C. & Greenaway, F. T. 2008. High-performance liquid chromatography/nano-electrospray ionization tandem mass spectrometry, two-dimensional difference in-gel electrophoresis and gene microarray identification of lymphatic metastasis-associated biomarkers. *Rapid Communications in Mass Spectrometry,* 22**,** 3172-8.

Llorach, R., Garcia-Aloy, M., Tulipani, S., Vazquez-Fresno, R. & Andres-Lacueva, C. 2012. Nutrimetabolomic strategies to develop new biomarkers of intake and health effects. *Journal of Agricultural and Food Chemistry,* 60**,** 8797-808.

Lloyd, A. J., Fave, G., Beckmann, M., Lin, W., Tailliart, K., Xie, L., Mathers, J. C. & Draper, J. 2011. Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *American Journal of Clinical Nutrition,* 94**,** 981-991.

Lommen, A. & Kools, H. J. 2012. MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. *Metabolomics,* 8**,** 719-726.

Loo, R. L., Chan, Q., Brown, I. J., Robertson, C. E., Stamler, J., Nicholson, J. K., Holmes, E., Elliott, P. & Grp, I. R. 2012. A Comparison of Self-Reported Analgesic Use and Detection of Urinary Ibuprofen and Acetaminophen Metabolites by Means of Metabonomics The INTERMAP Study. *American Journal of Epidemiology,* 175**,** 348-358.

Lu, W., Bennett, B. D. & Rabinowitz, J. D. 2008. Analytical strategies for LC-MS-based targeted metabolomics. *Journal of Chromatography B: Biomedical Sciences and Applications* 871**,** 236-42.

Ma, B., Zhang, Q., Wang, G. J., A, J. Y., Wu, D., Liu, Y., Cao, B., Liu, L. S., Hu, Y. Y., Wang, Y. L. & Zheng, Y. Y. 2011. GC-TOF/MS-based metabolomic profiling of estrogen deficiency-induced obesity in ovariectomized rats. *Acta Pharmacologica Sinica,* 32**,** 270-8.

Malherbe, G., Steel, H. C., Cassol, S., De Oliveira, T., Seebregts, C. J., Anderson, R., Cassol, E. & Rossouw, T. M. 2014. Circulating Biomarkers of Immune Activation Distinguish Viral Suppression from Nonsuppression in HAART-Treated Patients with Advanced HIV-1 Subtype C Infection. *Mediators of Inflammation,* 2014**,** 7.

Mamas, M., Dunn, W. B., Neyses, L. & Goodacre, R. 2011. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Archives of Toxicology,* 85**,** 5-17.

Manna, S. K., Patterson, A. D., Yang, Q., Krausz, K. W., Idle, J. R., Fornace, A. J. & Gonzalez, F. J. 2011. UPLC-MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the Ppara-null mouse model. *Journal of Proteome Research,* 10**,** 4120-33.

Mar, J. C., Kimura, Y., Schroder, K., Irvine, K. M., Hayashizaki, Y., Suzuki, H., Hume, D. & Quackenbush, J. 2009. Data-driven normalization strategies for high-throughput quantitative RT-PCR. *BMC Bioinformatics,* 10**,** 110-110.

March, R. E. 1997. An introduction to quadrupole ion trap mass spectrometry. *Journal of Mass Spectrometry,* 32**,** 351-369.

Marginean, I., Kelly, R. T., Prior, D. C., Lamarche, B. L., Tang, K. Q. & Smith, R. D. 2008. Analytical characterization of the electrospray ion source in the nanoflow regime. *Analytical Chemistry,* 80**,** 6573-6579.

Marginean, I., Tang, K., Smith, R. D. & Kelly, R. T. 2014. Picoelectrospray ionization mass spectrometry using narrow-bore chemically etched emitters. *Journal of the American Society for Mass Spectrometry,* 25**,** 30-36.

Marin, B., Thiebaut, R., Bucher, H. C., Rondeau, V., Costagliola, D., Dorrucci, M., Hamouda, O., Prins, M., Walker, S., Porter, K., Sabin, C. & Chene, G. 2009. Non-AIDS-defining deaths and immunodeficiency in the era of combination antiretroviral therapy. *AIDS,* 23**,** 1743-1752.

Marshall, A. G. 1985. Fourier transform ion cyclotron resonance mass spectrometry. *Accounts of Chemical Research,* 18**,** 316-322.

Marshall, A. G., Hendrickson, C. L. & Jackson, G. S. 1998. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrometry Reviews,* 17**,** 1-35.

Martinez, E., Gonzalez-Cordon, A., Ferrer, E., Domingo, P., Negredo, E., Gutierrez, F., Portilla, J., Curran, A., Podzamczer, D., Murillas, J., Bernardino, J. I., Santos, I., Carton, J. A., Peraire, J., Pich, J., Perez, I., Gatell, J. M. & Group, A. S. 2014. Early lipid changes with atazanavir/ritonavir or darunavir/ritonavir. *Hiv Medicine,* 15**,** 330-338.

Mattarucchi, E. & Guillou, C. 2011. Comment on "Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery". *Analytical Chemistry,* 83**,** 9719-9720.

Mattarucchi, E. & Guillou, C. 2012. Critical aspects of urine profiling for the selection of potential biomarkers using UPLC-TOF-MS. *Biomedical Chromatography,* 26**,** 512-517.

May, J. C., Goodwin, C. R. & Mclean, J. A. 2015. Ion mobility-mass spectrometry strategies for untargeted systems, synthetic, and chemical biology. *Current Opinion in Biotechnology,* 31**,** 117-121.

Mcrae, M., Rezk, N. L., Bridges, A. S., Corbett, A. H., Tien, H. C., Brouwer, K. L. & Kashuba, A. D. 2010. Plasma bile acid concentrations in patients with human immunodeficiency virus infection receiving protease inhibitor therapy: possible implications for hepatotoxicity. *Pharmacotherapy,* 30**,** 17-24.

Mcrae, M. P., Lowe, C. M., Tian, X., Bourdet, D. L., Ho, R. H., Leake, B. F., Kim, R. B., Brouwer, K. L. & Kashuba, A. D. 2006. Ritonavir, saquinavir, and efavirenz, but not nevirapine, inhibit bile acid transport in human and rat hepatocytes. *Journal of Pharmacology and Experimental Therapeutics,* 318**,** 1068-75.

Medina, S., Domínguez-Perles, R., Ferreres, F., Tomás-Barberán, F. A. & Gil-Izquierdo, Á. 2013a. The effects of the intake of plant foods on the human metabolome. *TrAC Trends in Analytical Chemistry,* 52**,** 88-99.

Medina, S., Ferreres, F., Garcia-Viguera, C., Horcajada, M. N., Orduna, J., Saviron, M., Zurek, G., Martinez-Sanz, J. M., Gil, J. I. & Gil-Izquierdo, A. 2013b. Non-targeted metabolomic approach reveals urinary metabolites linked to steroid biosynthesis pathway after ingestion of citrus juice. *Food Chemistry,* 136**,** 938-946.

Melzer, D., Rice, N. E., Lewis, C., Henley, W. E. & Galloway, T. S. 2010. Association of Urinary Bisphenol A Concentration with Heart Disease: Evidence from NHANES 2003/06. *Plos One,* 5.

Menéndez-Arias, L. 2002. Targeting HIV: antiretroviral therapy and development of drug resistance. *Trends in Pharmacological Sciences,* 23**,** 381-388.

Michell, A. W., Mosedale, D., Grainger, D. J. & Barker, R. A. 2008. Metabolomic analysis of urine and serum in Parkinson's disease. *Metabolomics,* 4**,** 191-201.

Miller, M. G. 2007. Environmental metabolomics: A SWOT analysis (strengths, weaknesses, opportunities, and threats). *Journal of Proteome Research,* 6**,** 540-545.

Mills, G. C. & Mills, J. S. 1985. Urinary excretion of methylthioadenosine in immunodeficient children. *Clinica Chimica Acta,* 147**,** 15-23.

Mills, G. C., Schmalstieg, F. C. & Goldblum, R. M. 1985. Urinary excretion of modified purines and nucleosides in immunodeficient children. *Biochemical Medicine,* 34**,** 37-51.

Moco, S., Bino, R. J., De Vos, R. C. H. & Vervoort, J. 2007. Metabolomics technologies and metabolite identification. *Trac-Trends in Analytical Chemistry,* 26**,** 855-866.

Morgan, R. E., Trauner, M., Van Staden, C. J., Lee, P. H., Ramachandran, B., Eschenberg, M., Afshari, C. A., Qualls, C. W., Lightfoot-Dunn, R. & Hamadeh, H. K. 2010. Interference with Bile Salt Export Pump Function Is a Susceptibility Factor for Human Liver Injury in Drug Development. *Toxicological Sciences,* 118**,** 485-500.

Morrens, B., Bruckers, L., Den Hond, E., Nelen, V., Schoeters, G., Baeyens, W., Van Larebeke, N., Keune, H., Bilau, M. & Loots, I. 2012. Social distribution of internal exposure to environmental pollution in Flemish adolescents. *International Journal of Hygiene and Environmental Health,* 215**,** 474-481.

Motokawa, M., Kobayashi, H., Ishizuka, N., Minakuchi, H., Nakanishi, K., Jinnai, H., Hosoya, K., Ikegami, T. & Tanaka, N. 2002. Monolithic silica columns with various skeleton sizes and through-pore sizes for capillary liquid chromatography. *Journal of Chromatography A,* 961**,** 53-63.

Munson, M. S. & Field, F.-H. 1966. Chemical ionization mass spectrometry. I. General introduction. *Journal of the American Chemical Society,* 88**,** 2621-2630.

Murphy, G. & Parry, J. V. 2008. Assays for the detection of recent infections with human immunodeficiency virus type 1. *Eurosurveillance,* 13**,** 18966.

Nahnsen, S., Bielow, C., Reinert, K. & Kohlbacher, O. 2013. Tools for Label-free Peptide Quantification. *Molecular & Cellular Proteomics,* 12**,** 549-556.

Neue, U. D., O'gara, J. E. & Mendez, A. 2006. Selectivity in reversed-phase separations - Influence of the stationary phase. *Journal of Chromatography A,* 1127**,** 161-174.

New, L. S. & Chan, E. C. Y. 2008. Evaluation of BEH C-18, BEH HILIC, and HSS T3 (C-18) column chemistries for the UPLC-MS-MS analysis of glutathione, glutathione disulfide, and ophthalmic acid in mouse liver and human plasma. *Journal of Chromatographic Science,* 46**,** 209-214.

Nguyen, D. T. T., Guillarme, D., Rudaz, S. & Veuthey, J. L. 2006. Fast analysis in liquid chromatography using small particle size and high pressure. *Journal of Separation Science,* 29**,** 1836-1848.

Nicholson, J. K. 2006. Global systems biology, personalized medicine and molecular epidemiology. *Molecular Systems Biology,* 2.

Nicholson, J. K. & Lindon, J. C. 2008. Systems biology - Metabonomics. *Nature,* 455**,** 1054-1056.

Nicholson, J. K., Lindon, J. C. & Holmes, E. 1999. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica,* 29**,** 1181-1189.

Nicholson, J. K. & Wilson, I. D. 2003. Understanding 'global' systems biology: Metabonomics and the continuum of metabolism. *Nature Reviews Drug Discovery,* 2**,** 668-676.

Nomura, H., Ueyama, J., Kondo, T., Saito, I., Murata, K., Iwata, T., Wakusawa, S. & Kamijima, M. 2013. Quantitation of neonicotinoid metabolites in human urine using GC-MS. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 941**,** 109-115.

Novakova, L. & Vlckova, H. 2009. A review of current trends and advances in modern bio-analytical methods: Chromatography and sample preparation. *Analytica Chimica Acta,* 656**,** 8-35.

O'sullivan, A., Gibney, M. J. & Brennan, L. 2011. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *American Journal of Clinical Nutrition,* 93**,** 314-321.

Paglia, G., Williams, J. P., Menikarachchi, L., Thompson, J. W., Tyldesley-Worster, R., Halldórsson, S., Rolfsson, O., Moseley, A., Grant, D., Langridge, J., Palsson, B. O. &

Astarita, G. 2014. Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Analytical Chemistry,* 86**,** 3985-3993.

Pantaleo, G., Graziosi, C. & Fauci, A. S. 1993a. The Immunopathogenesis of Human Immunodeficiency Virus Infection. *New England Journal of Medicine,* 328**,** 327-335.

Pantaleo, G., Graziosi, C. & Fauci, A. S. 1993b. New concepts in the immunopathogenesis of human immunodeficiency virus infection. *N Engl J Med,* 328**,** 327-35.

Pascal, S., Resnick, L., Barker, W. W., Loewenstein, D., Yoshii, F., Chang, J. Y., Boothe, T., Sheldon, J. & Duara, R. 1991. Metabolic asymmetries in asymptomatic HIV-1 seropositive subjects: relationship to disease onset and MRI findings. *Journal of Nuclear Medicine,* 32**,** 1725-9.

Pasikanti, K. K., Esuvaranathan, K., Hong, Y. J., Ho, P. C., Mahendran, R., Mani, L. R. N., Chiong, E. & Chan, E. C. Y. 2013. Urinary Metabotyping of Bladder Cancer Using Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry. *Journal of Proteome Research,* 12**,** 3865-3873.

Pasikanti, K. K., Ho, P. C. & Chan, E. C. Y. 2008. Development and validation of a gas chromatography/mass spectrometry metabonomic platform for the global profiling of urinary metabolites. *Rapid Communications in Mass Spectrometry,* 22**,** 2984-2992.

Pendyala, G., Want, E. J., Webb, W., Siuzdak, G. & Fox, H. S. 2007. Biomarkers for neuroAIDS: The widening scope of metabolomics. *Journal of Neuroimmune Pharmacology,* 2**,** 72-80.

Peng, J., Chen, Y.-T., Chen, C.-L. & Li, L. 2014. Development of a Universal Metabolome-Standard Method for Long-Term LC–MS Metabolome Profiling and Its Application for Bladder Cancer Urine-Metabolite-Biomarker Discovery. *Analytical Chemistry*.

Perrett, D. 2007. From 'protein' to the beginnings of clinical proteomics. *Proteomics Clinical Applications,* 1**,** 720-738.

Philippeos, C., Steffens, F. E. & Meyer, D. 2009. Comparative 1H NMR-based metabonomic analysis of HIV-1 sera. *Journal of Biomolecular NMR,* 44**,** 127-37.

Pineda, J. A., Palacios, R., Rivero, A., Abdel-Kader, L., Marquez, M., Cano, P., Mira, J. A., Torre-Cisneros, J., Lozano, F. & Santos, J. 2006. Low incidence of severe liver toxicity in patients receiving antiretroviral combinations including atazanavir. *Journal of Antimicrobial Chemotherapy,* 57**,** 1016-1017.

Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. 2010. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics,* 11**,** 395.

Pocock, G. & Richards, C. 2006. *Human Physiology The Basis of Medicine,* Oxford, Oxford University Press.

Portilla, D., Li, S., Nagothu, K. K., Megyesi, J., Kaissling, B., Schnackenberg, L., Safirstein, R. L. & Beger, R. D. 2006. Metabolomic study of cisplatin-induced nephrotoxicity. *Kidney International,* 69**,** 2194-2204.

Powers, R. 2009. NMR metabolomics and drug discovery. *Magnetic Resonance in Chemistry,* 47**,** S2-S11.

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., Greiner, R., Mcmanus, B., Newman, J. W., Goodfriend, T. & Wishart, D. S. 2011. The Human Serum Metabolome. *Plos One,* 6.

Pujos-Guillot, E., Hubert, J., Martin, J. F., Lyan, B., Quintana, M., Claude, S., Chabanas, B., Rothwell, J. A., Bennetau-Pelissero, C., Scalbert, A., Comte, B., Hercberg, S., Morand, C., Galan, P. & Manach, C. 2013. Mass Spectrometry-based Metabolomics for the Discovery of Biomarkers of Fruit and Vegetable Intake: Citrus Fruit as a Case Study. *Journal of Proteome Research,* 12**,** 1645-1659.

Qi, D., Brownridge, P., Xia, D., Mackay, K., Gonzalez-Galarza, F. F., Kenyani, J., Harman, V., Beynon, R. J. & Jones, A. R. 2012. A software toolkit and interface for performing stable isotope labeling and top3 quantification using Progenesis LC-MS. *Omics,* 16**,** 489-95.

Rabkin, J. G., Rabkin, R. & Wagner, G. J. 1997. Testosterone treatment of clinical hypogonadism in patients with HIV/AIDS. *International Journal of STD & AIDS,* 8**,** 537-545.

Rafiei, A. & Sleno, L. 2015. Comparison of peak-picking workflows for untargeted liquid chromatography/high-resolution mass spectrometry metabolomics data analysis. *Rapid Communications in Mass Spectrometry,* 29**,** 119-127.

Ramautar, R., Somsen, G. W. & De Jong, G. J. 2009. CE-MS in metabolomics. *Electrophoresis,* 30**,** 276-291.

Ramis, I., Rosello-Catafau, J., Gomez, G., Zabay, J. M., Fernandez Cruz, E. & Gelpi, E. 1991. Cyclooxygenase and lipoxygenase arachidonic acid metabolism by monocytes from human immune deficiency virus-infected drug users. *J Chromatogr,* 557**,** 507-13.

Rappaport, S. M. 2011. Implications of the exposome for exposure science. *Journal of Exposure Science and Environmental Epidemiology,* 21**,** 5-9.

Raterink, R.-J., Lindenburg, P. W., Vreeken, R. J., Ramautar, R. & Hankemeier, T. 2014. Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics. *TrAC Trends in Analytical Chemistry,* 61**,** 157-167.

Rekić, D., Clewe, O., Röshammar, D., Flamholc, L., Sönnerborg, A., Ormaasen, V., Gisslén, M., Äbelö, A. & Ashton, M. 2011. Bilirubin—A Potential Marker of Drug Exposure in Atazanavir-Based Antiretroviral Therapy. *The AAPS Journal,* 13**,** 598-605.

Rittweger, M. 2007. Clinical pharmacokinetics of darunavir. *Clinical pharmacokinetics,* 46**,** 739-756.

Rivero, A., Mira, J. A. & Pineda, J. A. 2007. Liver toxicity induced by non-nucleoside reverse transcriptase inhibitors. *Journal of Antimicrobial Chemotherapy,* 59**,** 342-346.

Robb, D. B., Covey, T. R. & Bruins, A. P. 2000. Atmospheric pressure photoionization: an ionization method for liquid chromatography-mass spectrometry. *Analytical Chemistry,* 72**,** 3653-3659.

Rochester, J. R. 2013. Bisphenol A and human health: A review of the literature. *Reproductive Toxicology,* 42**,** 132-155.

Rogeberg, M., Malerod, H., Roberg-Larsen, H., Aass, C. & Wilson, S. R. 2014. On-line solid phase extraction–liquid chromatography, with emphasis on modern bioanalysis and miniaturized systems. *Journal of Pharmaceutical and Biomedical Analysis,* 87**,** 120-129.

Rosner, M. H. & Okusa, M. D. 2006. Acute Kidney Injury Associated with Cardiac Surgery. *Clinical Journal of the American Society of Nephrology,* 1**,** 19-32.

Ryan, D., Robards, K., Prenzler, P. D. & Kendall, M. 2011. Recent and potential developments in the analysis of urine: A review. *Analytica Chimica Acta,* 684**,** 17-29.

Sabin, C. 2013. Do people with HIV infection have a normal life expectancy in the era of combination antiretroviral therapy? *BMC Medicine,* 11**,** 1-7.

Sadee, W. 2011. Genomics and personalized medicine. *International Journal of Pharmaceutics,* 415**,** 2-4.

Samuelsson, L. M., Forlin, L., Karlsson, G., Adolfsson-Eric, M. & Larsson, D. G. J. 2006. Using NMR metabolomics to identify responses of an environmental estrogen in blood plasma of fish. *Aquatic Toxicology,* 78**,** 341-349.

Sane, J., Heijman, T., Hogema, B., Koot, M., Van Veen, M., Götz, H., Fennema, J. & De Coul, E. O. 2014. Identifying recently acquired HIV infections among newly diagnosed men who have sex with men attending STI clinics in The Netherlands. *Sexually Transmitted Infections***,** doi: 10.1136/sextrans-2013-051420.

Satheeshkumar, N., Nisha, N., Sonali, N., Nirmal, J., Jain, G. K. & Spandana, V. 2012. Analytical Profiling of Bioactive Constituents from Herbal Products, using Metabolomics - A Review. *Natural Product Communications,* 7**,** 1111-1115.

Saude, E. J., Adamko, D., Rowe, B. H., Marrie, T. & Sykes, B. D. 2007. Variation of metabolites in normal human urine. *Metabolomics,* 3**,** 439-451.

Saude, E. J. & Sykes, B. D. 2007. Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics,* 3**,** 19-27.

Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T., Kristal, B. S., Van Ommen, B., Pujos-Guillot, E., Verheij, E., Wishart, D. & Wopereis, S. 2009. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics,* 5**,** 435-458.

Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L. O., Draper, J., Rappaport, S. M., Van Der Hooft, J. J. J. & Wishart, D. S. 2014. The food metabolome: a window over dietary exposure. *American Journal of Clinical Nutrition,* 99**,** 1286-1308.

Schlosser, A. & Volkmer-Engert, R. 2003. Volatile polydimethylcyclosiloxanes in the ambient laboratory air identified as source of extreme background signals in nanoelectrospray mass spectrometry. *Journal of Mass Spectrometry,* 38**,** 523-525.

Schmidt, A., Karas, M. & Dulcks, T. 2003. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? *Journal of the American Society for Mass Spectrometry,* 14**,** 492-500.

Schnackenberg, L. K., Sun, J., Pence, L. M., Bhattacharyya, S., Da Costa, G. G. & Beger, R. D. 2012. Metabolomics evaluation of hydroxyproline as a potential marker of melamine and cyanuric acid nephrotoxicity in male and female Fischer F344 rats. *Food and Chemical Toxicology,* 50**,** 3978-3983.

Schroijen, C., Baeyens, W., Schoeters, G., Den Hond, E., Koppen, G., Bruckers, L., Nelen, V., Van De Mieroop, E., Bilau, M., Covaci, A., Keune, H., Loots, I., Kleinjans, J., Dhooge, W. & Van Larebeke, N. 2008. Internal exposure to pollutants measured in blood and urine of Flemish adolescents in function of area of residence. *Chemosphere,* 71**,** 1317-1325.

Schwartz, J. C., Syka, J. E. P. & Jardine, I. 1991. High resolution on a quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry,* 2**,** 198-204.

Seeliger, E., Sendeski, M., Rihal, C. S. & Persson, P. B. 2012. *Contrast-induced kidney injury: mechanisms, risk factors, and prevention*.

Seli, E., Bruce, C., Botros, L., Henson, M., Roos, P., Judge, K., Hardarson, T., Ahlström, A., Harrison, P., Henman, M., Go, K., Acevedo, N., Siques, J., Tucker, M. & Sakkas, D. 2011. Receiver operating characteristic (ROC) analysis of day 5 morphology grading and metabolomic Viability Score on predicting implantation outcome. *Journal of Assisted Reproduction and Genetics,* 28**,** 137-144.

Serkova, N. J., Spratlin, J. L. & Eckhardt, S. G. 2007. NMR-based metabolomics: Translational application and treatment of cancer. *Current Opinion in Molecular Therapeutics,* 9**,** 572-585.

Sharp, P. M. & Hahn, B. H. 2011. Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine,* 1**,** 1-22.

Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. 1996. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Analytical Chemistry,* 68**,** 850-858.

Shockcor, J. P. & Holmes, E. 2002. Metabonomic applications in toxicity screening and disease diagnosis. *Current Topics in Medicinal Chemistry,* 2**,** 35-51.

Shrestha, B. & Vertes, A. 2014. High-Throughput Cell and Tissue Analysis with Enhanced Molecular Coverage by Laser Ablation Electrospray Ionization Mass Spectrometry Using Ion Mobility Separation. *Analytical Chemistry,* 86**,** 4308-4315.

Singh, R. K., Gupta, S., Dastidar, S. & Ray, A. 2010. Cysteinyl Leukotrienes and Their Receptors: Molecular and Functional Characteristics. *Pharmacology,* 85**,** 336-349.

Sitole, L. J., Williams, A. A. & Meyer, D. 2013. Metabonomic analysis of HIV-infected biofluids. *Molecular Biosystems,* 9**,** 18-28.

Skurtveit, S., Selmer, R., Tverdal, A. & Furu, K. 2008. The validity of self-reported prescription medication use among adolescents varied by therapeutic class. *Journal of Clinical Epidemiology,* 61**,** 714-717.

Slupsky, C. M., Steed, H., Wells, T. H., Dabbs, K., Schepansky, A., Capstick, V., Faught, W. & Sawyer, M. B. 2010. Urine Metabolite Analysis Offers Potential Early Diagnosis of Ovarian and Breast Cancers. *Clinical Cancer Research,* 16**,** 5835-5841.

Smith, C. A., Want, E. J., O'maille, G., Abagyan, R. & Siuzdak, G. 2006. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry,* 78**,** 779-787.

Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J. & Udseth, H. R. 1990. New developments in biochemical mass spectrometry: electrospray ionization. *Analytical Chemistry,* 62**,** 882-899.

Snyder, M. L. & Lichstein, H. C. 1940. Sodium azide as an inhibiting substance for gram negative bacteria. *Journal of Infectious Diseases,* 67**,** 113-115.

Solomon, R. & Dauerman, H. L. 2010. Contrast-Induced Acute Kidney Injury. *Circulation,* 122**,** 2451-2455.

Song, I.-S., Lee, D. Y., Shin, M.-H., Kim, H., Ahn, Y. G., Park, I., Kim, K. H., Kind, T., Shin, J.-G., Fiehn, O. & Liu, K.-H. 2012. Pharmacogenetics Meets Metabolomics: Discovery of Tryptophan as a New Endogenous OCT2 Substrate Related to Metformin Disposition. *Plos One,* 7**,** e36637.

Southam, A. D., Payne, T. G., Cooper, H. J., Arvanitis, T. N. & Viant, M. R. 2007. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Analytical Chemistry,* 79**,** 4595-4602.

Spagou, K., Wilson, I. D., Masson, P., Theodoridis, G., Raikos, N., Coen, M., Holmes, E., Lindon, J. C., Plumb, R. S., Nicholson, J. K. & Want, E. J. 2011. HILIC-UPLC-MS for Exploratory Urinary Metabolic Profiling in Toxicological Studies. *Analytical Chemistry,* 83**,** 382-390.

Sreekumar, A., Poisson, L. M., Rajendiran, T. M., Khan, A. P., Cao, Q., Yu, J. D., Laxman, B., Mehra, R., Lonigro, R. J., Li, Y., Nyati, M. K., Ahsan, A., Kalyana-Sundaram, S., Han, B., Cao, X. H., Byun, J., Omenn, G. S., Ghosh, D., Pennathur, S., Alexander, D. C., Berger, A., Shuster, J. R., Wei, J. T., Varambally, S., Beecher, C. & Chinnaiyan, A. M. 2009. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature,* 457**,** 910-914.

Sterz, K., Scherer, G. & Ecker, J. 2012. A simple and robust UPLC-SRM/MS method to quantify urinary eicosanoids. *Journal of Lipid Research,* 53**,** 1026-1036.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W. M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reily, M. D., Thaden, J. J. & Viant, M. R. 2007. Proposed minimum reporting standards for chemical analysis. *Metabolomics,* 3**,** 211-221.

Sun, J., Schnackenberg, L. K., Holland, R. D., Schmitt, T. C., Cantor, G. H., Dragan, Y. R. & Beger, R. D. 2008. Metabonomics evaluation of urine from rats given acute and chronic doses of acetaminophen using NMR and UPLC/MS. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 871**,** 328-340.

Sun, J. C., Bhattacharyya, S., Schnackenberg, L. K., Pence, L., Ando, Y., Zhang, J., Stewart, S., Rosenzweig, B., Rouse, R., Portilla, D. & Beger, R. D. 2012. Discovery of early urinary biomarkers in preclinical study of gentamicin-induced kidney injury and recovery in rats. *Metabolomics,* 8**,** 1181-1193.

Swartz, M. E. 2005. UPLC (TM): An introduction and review. *Journal of Liquid Chromatography & Related Technologies,* 28**,** 1253-1263.

Tautenhahn, R., Cho, K., Uritboonthai, W., Zhu, Z. J., Patti, G. J. & Siuzdak, G. 2012a. An accelerated workflow for untargeted metabolomics using the METLIN database. *Nature Biotechnology,* 30**,** 826-828.

Tautenhahn, R., Patti, G. J., Rinehart, D. & Siuzdak, G. 2012b. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Analytical Chemistry,* 84**,** 5035-5039.

Ter Heine, R., Hillebrand, M. J., Rosing, H., Van Gorp, E. C., Mulder, J. W., Beijnen, J. H. & Huitema, A. D. 2009. Identification and profiling of circulating metabolites of atazanavir, a HIV protease inhibitor. *Drug Metabolism and Disposition,* 37**,** 1826-40.

Theodoridis, G., Gika, H. G. & Wilson, I. D. 2008. LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. *Trac-Trends in Analytical Chemistry,* 27**,** 251-260.

Theodoridis, G. A., Gika, H. G., Want, E. J. & Wilson, I. D. 2012. Liquid chromatography-mass spectrometry based global metabolite profiling: A review. *Analytica Chimica Acta,* 711**,** 7-16.

Tomaka, F., Lefebvre, E., Sekar, V., Van Baelen, B., Vangeneugden, T., Vandevoorde, A. & Diego Miralles, G. 2009. Effects of ritonavir-boosted darunavir vs. ritonavir-boosted atazanavir on lipid and glucose parameters in HIV-negative, healthy volunteers. *Hiv Medicine,* 10**,** 318-327.

Trupp, M., Zhu, H., Wikoff, W. R., Baillie, R. A., Zeng, Z.-B., Karp, P. D., Fiehn, O., Krauss, R. M. & Kaddurah-Daouk, R. 2012. Metabolomics Reveals Amino Acids Contribute to Variation in Response to Simvastatin Treatment. *Plos One,* 7**,** e38386.

Ueyama, J., Saito, I., Kondo, T., Taki, T., Kimata, A., Saito, S., Ito, Y., Murata, K., Iwata, T., Gotoh, M., Shibata, E., Wakusawa, S. & Kamijima, M. 2012. Urinary concentrations of organophosphorus insecticide metabolites in Japanese workers. *Chemosphere,* 87**,** 1403-1409.

Van De Meent, M. H. M. & De Jong, G. J. 2011. Novel liquid-chromatography columns for proteomics research. *TrAC Trends in Analytical Chemistry,* 30**,** 1809-1818.

Van Den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. & Van Der Werf, M. J. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics,* 7.

Vela-Soria, F., Jimenez-Diaz, I., Rodriguez-Gomez, R., Zafra-Gomez, A., Ballesteros, O., Navalon, A., Vilchez, J. L., Fernandez, M. F. & Olea, N. 2011. Determination of benzophenones in human placental tissue samples by liquid chromatography-tandem mass spectrometry. *Talanta,* 85**,** 1848-55.

Vergeynst, L., Van Langenhove, H., Joos, P. & Demeestere, K. 2013. Accurate mass determination, quantification and determination of detection limits in liquid chromatography-high-resolution time-of-flight mass spectrometry: Challenges and practical solutions. *Analytica Chimica Acta,* 789**,** 74-82.

Vermeir, M., Lachau-Durand, S., Mannens, G., Cuyckens, F., Van Hoof, B. & Raoof, A. 2009. Absorption, Metabolism, and Excretion of Darunavir, a New Protease Inhibitor, Administered Alone and with Low-Dose Ritonavir in Healthy Subjects. *Drug Metabolism and Disposition,* 37**,** 809-820.

Veselkov, K. A., Vingara, L. K., Masson, P., Robinette, S. L., Want, E., Li, J. V., Barton, R. H., Boursier-Neyret, C., Walther, B., Ebbels, T. M., Pelczer, I., Holmes, E., Lindon, J. C. & Nicholson, J. K. 2011a. Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery. *Analytical Chemistry,* 83**,** 5864-5872.

Veselkov, K. A., Vingara, L. K., Masson, P., Robinette, S. L., Want, E., Li, J. V., Barton, R. H., Boursier-Neyret, C., Walther, B., Ebbels, T. M., Pelczer, I., Holmes, E., Lindon, J. C. & Nicholson, J. K. 2011b. Response to Comment on "Optimized Preprocessing of Ultra-

Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery". *Analytical Chemistry,* 83**,** 9721-9722.

Viant, M. R., Rosenblum, E. S. & Tjeerdema, R. S. 2003. NMR-based metabolomics: A powerful approach for characterizing the effects of environmental stressors on organism health. *Environmental Science & Technology,* 37**,** 4982-4989.

Vinayavekhin, N., Homan, E. A. & Saghatelian, A. 2010. Exploring Disease through Metabolomics. *ACS Chemical Biology,* 5**,** 91-103.

Wagner, G., Rabkin, J. G. & Rabkin, R. 1995. Illness Stage, Concurrent Medications, and Other Correlates of Low Testosterone in Men with HIV Illness. *JAIDS Journal of Acquired Immune Deficiency Syndromes,* 8**,** 204-207.

Walsh, M. C., Brennan, L., Malthouse, J. P. G., Roche, H. M. & Gibney, M. J. 2006. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *American Journal of Clinical Nutrition,* 84**,** 531-539.

Wang, C., Feng, R. N., Sun, D. J., Li, Y., Bi, X. X. & Sun, C. H. 2011. Metabolic profiling of urine in young obese men using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC/Q-TOF MS). *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 879**,** 2871-2876.

Wang, C., Zuo, Y., Vinson, J. A. & Deng, Y. 2012a. Absorption and excretion of cranberry-derived phenolics in humans. *Food Chemistry,* 132**,** 1420-1428.

Wang, F., Dong, J., Jiang, X., Ye, M. & Zou, H. 2007. Capillary Trap Column with Strong Cation-Exchange Monolith for Automated Shotgun Proteome Analysis. *Analytical Chemistry,* 79**,** 6599-6606.

Wang, S., Xu, F. & Demirci, U. 2010a. Advances in developing HIV-1 viral load assays for resource-limited settings. *Biotechnology Advances,* 28**,** 770-781.

Wang, W., Feng, B., Li, X., Yin, P., Gao, P., Zhao, X., Lu, X., Zheng, M. & Xu, G. 2010b. Urinary metabolic profiling of colorectal carcinoma based on online affinity solid phase extraction-high performance liquid chromatography and ultra performance liquid chromatography-mass spectrometry. *Molecular Biosystems,* 6**,** 1947-1955.

Wang, X. N., Wang, X. Y., Xie, G. X., Zhou, M. M., Yu, H., Lin, Y., Du, G. L., Luo, G. A., Jia, W. & Liu, P. 2012b. Urinary Metabolite Variation Is Associated with Pathological Progression of the Post-Hepatitis B Cirrhosis Patients. *Journal of Proteome Research,* 11**,** 3838-3847.

Want, E. & Masson, P. 2011. Processing and Analysis of GC/LC-MS-Based Metabolomics Data. *In:* METZ, T. O. (ed.) *Metabolic Profiling.* Humana Press.

Want, E. J., Cravatt, B. F. & Siuzdak, G. 2005. The expanding role of mass spectrometry in metabolite profiling and characterization. *Chembiochem,* 6**,** 1941-1951.

Want, E. J., Nordstrom, A., Morita, H. & Siuzdak, G. 2007. From exogenous to endogenous: the inevitable imprint of mass spectrometry in metabolomics. *J Proteome Res,* 6**,** 459-68.

Want, E. J., O'maille, G., Smith, C. A., Brandon, T. R., Uritboonthai, W., Qin, C., Trauger, S. A. & Siuzdak, G. 2006. Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Analytical Chemistry,* 78**,** 743-752.

Want, E. J., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, R. S., Shockcor, J., Holmes, E. & Nicholson, J. K. 2010. Global metabolic profiling procedures for urine using UPLC-MS. *Nature Protocols,* 5**,** 1005-1018.

Warrack, B. M., Hnatyshyn, S., Ott, K. H., Reily, M. D., Sanders, M., Zhang, H. Y. & Drexler, D. M. 2009. Normalization strategies for metabonomic analysis of urine samples. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 877**,** 547-552.

Watson, D. G. 2013. A Rough Guide to Metabolite Identification Using High Resolution Liquid Chromatography Mass Spectrometry in Metabolomic Profiling in Metazoans. *Computational and Structural Biotechnology Journal,* 4**,** e201301005.

Waybright, T. J., Van, Q. N., Muschik, G. M., Conrads, T. P., Veenstra, T. D. & Issaq, H. J. 2006. LC-MS in metabonomics: Optimization of experimental conditions for the analysis of metabolites in human urine. *Journal of Liquid Chromatography & Related Technologies,* 29**,** 2475-2497.

Weiss, R. H. & Kim, K. 2012. Metabolomics in the study of kidney diseases. *Nature Reviews Nephrology,* 8**,** 22-33.

Wen, H., Yang, H. J., Choi, M. J., Kwon, H. N., Kim, M. A., Hong, S. S. & Park, S. 2011. Identification of Urinary Biomarkers Related to Cisplatin-Induced Acute Renal Toxicity Using NMR-Based Metabolomics. *Biomolecules & Therapeutics,* 19**,** 38-44.

Werner, E., Croixmarie, V., Umbdenstock, T., Ezan, E., Chaminade, P., Tabet, J. C. & Junot, C. 2008. Mass spectrometry-based metabolomics: Accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution. *Analytical Chemistry,* 80**,** 4918-4932.

West, S. L., Savitz, D. A., Koch, G., Strom, B. L., Guess, H. A. & Hartzema, A. 1995. Recall Accuracy for Prescription Medications: Self-report Compared with Database Information. *American Journal of Epidemiology,* 142**,** 1103-1112.

Westerhuis, J. A., Van Velzen, E. J. J., Hoefsloot, H. C. J. & Smilde, A. K. 2010. Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics,* 6**,** 119-128.

Wickremsinhe, E. R., Singh, G., Ackermann, B. L., Gillespie, T. A. & Chaudhary, A. K. 2006. A review of nanoelectrospray ionization applications for drug metabolism and pharmacokinetics. *Current Drug Metabolism,* 7**,** 913-28.

Wiklund, S., Johansson, E., Sjostrom, L., Mellerowicz, E. J., Edlund, U., Shockcor, J. P., Gottfries, J., Moritz, T. & Trygg, J. 2008. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Analytical Chemistry,* 80**,** 115-122.

Wikoff, W. R., Pendyala, G., Siuzdak, G. & Fox, H. S. 2008. Metabolomic analysis of the cerebrospinal fluid reveals changes in phospholipase expression in the CNS of SIV-infected macaques. *Journal of Clinical Investigation,* 118**,** 2661-2669.

Wilcox, C. M., Cryer, B. & Triadafilopoulos, G. 2005. Patterns of use and public perception of over-the-counter pain relievers: focus on nonsteroidal antiinflammatory drugs. *The Journal of Rheumatology,* 32**,** 2218-2224.

Wild, C. P. 2012. The exposome: from concept to utility. *International Journal of Epidemiology,* 41**,** 24-32.

Williams, A., Koekemoer, G., Lindeque, Z., Reinecke, C. & Meyer, D. 2012. Qualitative serum organic acid profiles of HIV-infected individuals not on antiretroviral treatment. *Metabolomics,* 8**,** 804-818.

Williams, I., Churchill, D., Anderson, J., Boffito, M., Bower, M., Cairns, G., Cwynarski, K., Edwards, S., Fidler, S., Fisher, M., Freedman, A., Geretti, A. M., Gilleece, Y., Horne, R., Johnson, M., Khoo, S., Leen, C., Marshall, N., Nelson, M., Orkin, C., Paton, N., Phillips, A., Post, F., Pozniak, A., Sabin, C., Trevelion, R., Ustianowski, A., Walsh, J., Waters, L., Wilkins, E., Winston, A., Youle, M. & Writing, G. 2014. Special Issue: British HIV Association guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2012 (2013 update). *HIV Medicine,* 15**,** 1-85.

Wilm, M. & Mann, M. 1996. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry,* 68**,** 1-8.

Wilson, I. D., Plumb, R., Granger, J., Major, H., Williams, R. & Lenz, E. A. 2005. HPLC-MS-based methods for the study of metabonomics. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 817**,** 67-76.

Wishart, D. S. 2008. Metabolomics: applications to food science and nutrition research. *Trends in Food Science & Technology,* 19**,** 482-493.

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y. F., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J. G., Liu, P., Yallou, F.,

Bjorndahl, T., Perez-Pineiro, R., Eisner, R., Allen, F., Neveu, V., Greiner, R. & Scalbert, A. 2013. HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Research,* 41**,** D801-D807.

Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J. G., Jia, L., Cruz, J. A., Lim, E., Sobsey, C. A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y. P., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H. J. & Forsythe, I. 2009. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Research,* 37**,** D603-D610.

Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M. A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D. D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G. E., Macinnis, G. D., Weljie, A. M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B. D., Vogel, H. J. & Querengesser, L. 2007. HMDB: the human metabolome database. *Nucleic Acids Research,* 35**,** D521-D526.

Wittmann, B. M., Stirdivant, S. M., Mitchell, M. W., Wulff, J. E., Mcdunn, J. E., Li, Z., Dennis-Barrie, A., Neri, B. P., Milburn, M. V., Lotan, Y. & Wolfert, R. L. 2014. Bladder Cancer Biomarker Discovery Using Global Metabolomic Profiling of Urine. *Plos One,* 9**,** e115870.

Wood, A. J. & Flexner, C. 1998. HIV-protease inhibitors. *New England Journal of Medicine,* 338**,** 1281-1293.

Wu, H., Xue, R., Dong, L., Liu, T., Deng, C., Zeng, H. & Shen, X. 2009. Metabolomic profiling of human urine in hepatocellular carcinoma patients using gas chromatography/mass spectrometry. *Analytica Chimica Acta,* 648**,** 98-104.

Xie, G. X., Li, X., Li, H. K. & Jia, W. 2013. Toward Personalized Nutrition: Comprehensive Phytoprofiling and Metabotyping. *Journal of Proteome Research,* 12**,** 1547-1559.

Yamashita, M. & Fenn, J. B. 1984. Electrospray ion source. Another variation on the free-jet theme. *Journal of Physical Chemistry,* 88**,** 4451-4459.

Yang, Q., Shi, X., Wang, Y., Wang, W., He, H., Lu, X. & Xu, G. 2010. Urinary metabonomic study of lung cancer by a fully automatic hyphenated hydrophilic interaction/RPLC-MS system. *Journal of Separation Science,* 33**,** 1495-1503.

Ye, X. Y., Zhou, X. L., Furr, J., Ahn, K. C., Hammock, B. D., Gray, E. L. & Calafat, A. M. 2011. Biomarkers of exposure to triclocarban in urine and serum. *Toxicology,* 286**,** 69-74.

Yin, P. & Xu, G. 2014. Current state-of-the-art of nontargeted metabolomics based on liquid chromatography–mass spectrometry with special emphasis in clinical applications. *Journal of Chromatography A,* 1374**,** 1-13.

Yoshida, T. 2004. Peptide separation by Hydrophilic-Interaction Chromatography: a review. *Journal of Biochemical and Biophysical Methods,* 60**,** 265-280.

Yu, T., Park, Y., Li, S. & Jones, D. P. 2013. Hybrid feature detection and information accumulation using high-resolution LC- MS metabolomics data. *Journal of Proteome Research,* 12**,** 1419-1427.

Zacharias, H. U., Schley, G., Hochrein, J., Klein, M. S., Koberle, C., Eckardt, K. U., Willam, C., Oefner, P. J. & Gronwald, W. 2013. Analysis of human urine reveals metabolic changes related to the development of acute kidney injury following cardiac surgery. *Metabolomics,* 9**,** 697-707.

Zamora-Ros, R., Rabassa, M., Cherubini, A., Urpi-Sarda, M., Llorach, R., Bandinelli, S., Ferrucci, L. & Andres-Lacueva, C. 2011. Comparison of 24-h volume and creatinine-corrected total urinary polyphenol as a biomarker of total dietary polyphenols in the Invecchiare InCHIANTI study. *Analytica Chimica Acta,* 704**,** 110-115.

Zeldin, R. K. & Petruschke, R. A. 2004. Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients. *Journal of Antimicrobial Chemotherapy,* 53**,** 4-9.

Zhang, A., Zhou, X., Zhao, H., Guan, Y., Zhou, S., Yan, G.-L., Ma, Z., Liu, Q. & Wang, X. 2014. Rapidly improved determination of metabolites from biological data sets using the high-efficient TransOmics tool. *Molecular Biosystems,* 10**,** 2160-2165.

Zhang, A. H., Sun, H., Wang, P., Han, Y. & Wang, X. J. 2012a. Modern analytical techniques in metabolomics analysis. *Analyst,* 137**,** 293-300.

Zhang, D., Chando, T. J., Everett, D. W., Patten, C. J., Dehal, S. S. & Humphreys, W. G. 2005. In vitro inhibition of UDP glucuronosyltransferases by atazanavir and other HIV protease inhibitors and the relationship of this property to in vivo bilirubin glucuronidation. *Drug Metabolism and Disposition,* 33**,** 1729-1739.

Zhang, F., Jia, Z., Gao, P., Kong, H., Li, X., Lu, X., Wu, Y. & Xu, G. 2010. Metabonomics study of urine and plasma in depression and excess fatigue rats by ultra fast liquid chromatography coupled with ion trap-time of flight mass spectrometry. *Molecular Biosystems,* 6**,** 852-61.

Zhang, F. X., Jia, Z. H., Gao, P., Kong, H. W., Li, X., Chen, J., Yang, Q., Yin, P. Y., Wang, J. S., Lu, X., Li, F. M., Wu, Y. L. & Xu, G. W. 2009. Metabonomics study of atherosclerosis rats by ultra fast liquid chromatography coupled with ion trap-time of flight mass spectrometry. *Talanta,* 79**,** 836-844.

Zhang, T., Creek, D. J., Barrett, M. P., Blackburn, G. & Watson, D. G. 2012b. Evaluation of Coupling Reversed Phase, Aqueous Normal Phase, and Hydrophilic Interaction Liquid Chromatography with Orbitrap Mass Spectrometry for Metabolomic Studies of Human Urine. *Analytical Chemistry,* 84**,** 1994-2001.

Zhao, X. J., Fritsche, J., Wang, J. S., Chen, J., Rittig, K., Schmitt-Kopplin, P., Fritsche, A., Haring, H. U., Schleicher, E. D., Xu, G. W. & Lehmann, R. 2010. Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. *Metabolomics,* 6**,** 362-374.

Zheng, J.-H., Guida, L. A., Rower, C., Castillo-Mancilla, J., Meditz, A., Klein, B., Kerr, B. J., Langness, J., Bushman, L. & Kiser, J. 2014. Quantitation of tenofovir and emtricitabine in dried blood spots (DBS) with LC–MS/MS. *Journal of Pharmaceutical and Biomedical Analysis,* 88**,** 144-151.

Zheng, P., Chen, J. J., Huang, T., Wang, M. J., Wang, Y., Dong, M. X., Huang, Y. J., Zhou, L. K. & Xie, P. 2013. A Novel Urinary Metabolite Signature for Diagnosing Major Depressive Disorder. *Journal of Proteome Research,* 12**,** 5904-5911.

Zhou, M., Politis, A., Davies, R. B., Liko, I., Wu, K.-J., Stewart, A. G., Stock, D. & Robinson, C. V. 2014. Ion mobility–mass spectrometry of a rotary ATPase reveals ATP-induced reduction in conformational flexibility. *Nature Chemistry,* 6**,** 208-215.

Zubarev, R. A. & Makarov, A. 2013. Orbitrap Mass Spectrometry. *Analytical Chemistry,* 85**,** 5288-5296.